Artificial Agents as Leaders and Followers: An Integrative Role-Congruity Review

Christoph Daldrop[1], Annabelle Ludwig[1], and Claudia Buengeler[1,2]

[1]Department of Human Resource Management and Organization, Kiel University, Germany

[2]Amsterdam Business School, University of Amsterdam, The Netherlands

Author Note

Christoph Daldrop: https://orcid.org/0000-0003-3125-2035

Claudia Buengeler: https://orcid.org/0000-0003-1276-2232

Publication Note

This Author Accepted Manuscript corresponds in text to the published version, including the corrections made at the proof stage; layout and pagination differ from the version of record. It is archived on MACAU, the open access publication service of Kiel University, at https://doi.org/10.38071/2026-01499-0 under a CC BY-NC-ND 4.0 license.

Please cite the published version: Daldrop, C., Ludwig, A., & Buengeler, C. (2026). Artificial agents as leaders and followers: An integrative role-congruity review. *Group & Organization Management*. https://doi.org/10.1177/10596011261476917

Correspondence concerning this article should be addressed to Christoph Daldrop, Department of Human Resource Management and Organization, Faculty of Business, Economics and Social Sciences, Kiel University, Boschstr. 1, 24118 Kiel, Germany. E-mail: daldrop@bwl.uni-kiel.de

Abstract

Artificial agents such as algorithms, virtual agents, and robots increasingly occupy organizational roles as leaders, followers, and partners. However, evidence on how people evaluate these nonhuman role occupants is mixed. We systematically review 107 empirical studies across management, psychology, engineering, and computer science through a role-congruity lens. A consistent pattern emerges: Artificial agents in leader roles face role incongruity, driven largely by perceived deficits in morality (e.g., fairness, trustworthiness). Agents in follower roles are evaluated as largely congruent with role expectations. Agents in partner and teammate roles show conditional congruity that depends on task context and design-based social cues (e.g., appearance, voice). We integrate these patterns into the Artificial Agent Role Congruity (AARC) model, a conceptual model that extends role congruity theory to artificial agents: perceived role (in)congruity occurs when observers compare role prescriptions with the attributes they ascribe to an agent from its design-based cues along a task dimension (competence, assertiveness) and a social-emotional dimension (warmth, morality). Based on this synthesis, we outline theoretical extensions, an agenda for future research, and guidance for managers and designers deploying artificial agents in organizational roles.

Keywords: role incongruity, algorithmic management, robot leadership, human-AI collaboration, anthropomorphism, mind-role fit

Artificial Agents as Leaders and Followers: An Integrative Role-Congruity Review

> "I like robots. I like them considerably better than I do human beings. If a robot can be created capable of being a civil executive, I think he'd make the best one possible. By the Laws of Robotics, he'd be incapable of harming humans, incapable of tyranny, of corruption, of stupidity, of prejudice. […]" —Dr. Susan Calvin.
>
> *Isaac Asimov, I, Robot (1950).*

This quote from Asimov's science-fiction short story "Evidence" (Asimov, 1950) foreshadows a phenomenon that has become increasingly salient in organizations: the emergence of robots and artificial intelligence systems as managerial actors (e.g., Tsai, Marshall, et al., 2022; Wesche & Sonderegger, 2019). On digital platforms, algorithmic systems already allocate work, monitor performance, and issue feedback (e.g., Curchod et al., 2020; Rahman, 2021). Recent EU-representative worker surveys show that about one in five workers is already managed by software: between 16% and 27% report that algorithmic systems allocate tasks or monitor performance (Riso et al., 2025). Further, organizations are beginning to experiment with embodied forms such as virtual agents, including digital twins of executives that can attend meetings or respond to employee inquiries on their behalf (Boyle, 2025). However, hopes that algorithmic or robotic managers can deliver objective and bias-free decisions (e.g., Chamorro-Premuzic & Ahmetoglu, 2016) exist alongside perceptions that such nonhuman agents lack relevant social-emotional capacities such as empathy, care, and emotional understanding (e.g., Yam et al., 2025). This reveals a tension between nonhuman agents occupying leader roles and the human-centric nature of leadership (Tsai, Marshall, et al., 2022).

Empirical findings mirror this tension by revealing considerable inconsistencies: Perceptions can shift depending on the ascribed *organizational role*. For example, an identical chatbot labeled an "AI-based manager" was evaluated as more likable and knowledgeable than

when framed as an "AI representative" (Jeon, 2022). Adding to the complexity, perceptions also vary by *design*: robots high in human likeness can increase perceptions of warmth and acceptance (e.g., Wolf & Stock-Homburg, 2026), yet also be viewed as more abusive (e.g., Yam et al., 2022). Some studies show that algorithmic or robotic leaders are perceived as competent, fair, or even preferable to human leaders (e.g., Höddinghaus et al., 2021; Wolf & Stock-Homburg, 2026), whereas others find that they evoke discomfort, distrust, or illegitimacy (e.g., Jago et al., 2024; McGuire & De Cremer, 2023).

This pattern of mixed reactions to the same artificial agent across roles and designs points to a common evaluative mechanism: the perceived fit, or lack of fit, between what a role is seen to require and what the agent appears to offer. Role congruity theory (Eagly & Karau, 2002) explains this mechanism by comparing attributes that are *prescribed* by a professional role (e.g., what a leader should be like) and attributes that are *ascribed* on the basis of social categories such as gender or age (Eagly, 1987; Eagly & Wood, 2012; Koenig & Eagly, 2014). When prescribed role expectations and ascribed attributes are perceived as congruent, targets are seen as appropriate and legitimate; when they are perceived as incongruent, observers are more likely to respond with bias or resistance (del Carmen Triana et al., 2024; Eagly & Karau, 2002).

As algorithmic systems and robots are perceived as social actors in organizations (e.g., Anthony et al., 2023), we extend this logic of role (in)congruity to nonhuman agents. We define these *artificial agents* as machine-based systems that operate with varying autonomy, adapt after deployment, and generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments (European Union, 2024). Whereas humans in organizational roles are evaluated through social categories that shape perceived role incongruity, artificial agents lack such category memberships. Instead, we argue that observers

rely on design-based social cues, such as embodiment (e.g., algorithm vs. robot) and verbal and non-verbal behavior (e.g., voice, conversational style, name), to infer an agent's attributes (e.g., Blut et al., 2021; Glikson & Woolley, 2020). When the attributes inferred from these cues do not fit those prescribed by the role, observers perceive role incongruity, shaping how they evaluate, respond to, and adapt to artificial agents in organizational roles (del Carmen Triana et al., 2024). This lack-of-fit mechanism clarifies why the same artificial agent may be evaluated differently across roles and designs.

Research has typically examined roles and design cues separately rather than together. Work on artificial agents in organizations specified how these agents can occupy leader, partner, and follower roles (Tsai, Marshall, et al., 2022; Wesche & Sonderegger, 2019) and, in particular, when they can function as teammates in human–agent teams (e.g., Berretta et al., 2023; Bezrukova et al., 2023; Lyons et al., 2021; Seeber et al., 2020; Wolf & Stock-Homburg, 2023; Wynne & Lyons, 2018). Design-focused reviews show that appearance and verbal and non-verbal behavioral cues evoke perceptions of human likeness and mind perceptions that increase perceived competence and warmth and shape acceptance (e.g., Blut et al., 2021; Liew & Tan, 2021; Wirtz et al., 2018). Trust has emerged as one focal outcome of these evaluations, and trust-focused reviews highlight performance, transparency, context, and mind perception as central antecedents of trust in artificial teammates (e.g., Benk et al., 2025; Dang & Li, 2026; Glikson & Woolley, 2020; Hancock et al., 2021; Vanneste & Puranam, 2025; Wildman et al., 2024). Yet, roles and design cues have rarely been considered together, with only a few exceptions in preliminary work on machine replacement that adopt a narrower mind-role fit perspective (Dong et al., 2026; Yam et al., 2025). We therefore focus on role incongruity as the underlying

mechanism and combine organizational role expectations with design-based social cues to explain when and why people accept or resist artificial agents as leaders, followers, and partners.

In this integrative review, we map and explain inconsistencies in prior research on artificial agents in various organizational roles by uncovering the mechanism through which observers evaluate artificial agents as legitimate or illegitimate occupants of leader, follower, or partner roles. To this end, we systematically review 107 empirical studies across management, psychology, engineering, and computer science, apply role incongruity as the core explanatory lens (Eagly & Karau, 2002), and later integrate the recurring patterns into a conceptual model. When prescribed role expectations and ascribed attributes are perceived as congruent, artificial agents are judged as legitimate role occupants; when they are perceived as incongruent, illegitimacy perceptions dominate. We further organize the breadth of indicators used in (in)congruity evaluations into two fundamental dimensions of social perception: a task dimension capturing competence and assertiveness and a social-emotional dimension comprising warmth and morality (Abele et al., 2021). We focus on these dimensions because they summarize the criteria by which leader and follower roles are evaluated (e.g., Benson et al., 2024; Nichols et al., 2026).

Following a human-centric perspective on emerging technologies (e.g., Parker et al., 2026), we place observers' perceptions at the center of the analysis: Whether an artificial agent can function in an organizational role depends both on its capabilities *and* on the people who work with it. With our review, we contribute to this understanding in three ways. First, we organize and integrate a large and fragmented body of evidence spanning management, psychology, engineering, and computer science, and show that a role-incongruity lens extracts coherent patterns from findings that had appeared inconsistent: competent-yet-cold evaluations of agent

leaders, relative congruity for agent followers, and conditional acceptance of agent partners and teammates. Second, weaving in the recurring patterns identified in our review, we extend role congruity theory to artificial agents through the Artificial Agent Role Congruity (AARC) model, explaining how these differing evaluations emerge from the fit or misfit between attributes inferred from design-based cues and attributes prescribed for organizational roles. This clarifies that established role mechanisms apply when the role occupant is nonhuman, answering calls to ground human–AI research in established psychological theory (Bigman et al., 2026). Third, we set an AARC-based agenda for future research on role-design fit, spanning theoretical, methodological, and practical questions.

Conceptual Background

In this integrative review, we combine role congruity theory (del Carmen Triana et al., 2024; Eagly & Karau, 2002) with research on human–robot and human–agent collaboration (e.g., Tsai, Marshall, et al., 2022; Yam et al., 2025) into an evaluative lens for how observers perceive artificial agents in organizational roles. We organize our review around three elements: (1) organizational roles as prescriptive expectations (leader, follower, partner), (2) design-based features as social cues from which observers ascribe attributes, and (3) the comparison between prescribed role expectations and ascribed attributes underlying perceived role (in)congruity.

Role Congruity as Core Mechanism

Within organizations, actors occupy roles that carry distinct functions and expectations. Role theory conceptualizes these expectations as socially shared sets of norms that define appropriate attributes and behavior for occupants of specific positions (Anglin et al., 2022; Biddle, 1986). As a consequence, roles provide a reference for evaluation; observers compare what the organizational role prescribes with how they perceive a target. Role congruity theory

(Eagly & Karau, 2002) formalizes this evaluative mechanism as a comparison between two sets of expectations: the attributes prescribed by a functional role, such as a leader or follower role, and the attributes ascribed to a target based on a salient social category (e.g., gender; Eagly & Wood, 2012; Koenig & Eagly, 2014). When these attributes are *incongruent*, observers perceive or anticipate a *lack of fit* and respond with bias or resistance (Eagly & Karau, 2002; Heilman, 1983). For example, when leader roles emphasize assertiveness and authority, individuals who are instead stereotyped as more communal and less agentic may be evaluated as less congruent with these roles (Koenig et al., 2011).

We organize perceived (in)congruity around two fundamental dimensions of social perception: a *task dimension* capturing competence and assertiveness, and a *social-emotional dimension* capturing warmth and morality (Abele et al., 2016, 2021). Competence and assertiveness both concern accomplishing work and exerting influence, whereas warmth and morality both concern relating to others, and they relate more strongly within than between the two dimensions (Koch et al., 2024; Yzerbyt et al., 2025). Observers perceive a target along these dimensions and compare this perception with what the role prescribes; (in)congruity is the fit or misfit between the two.

Further, management research distinguishes three classes of role-incongruity effects: target evaluations, target outcomes, and target adaptation (del Carmen Triana et al., 2024). We use these *evaluations*, such as perceived effectiveness and favorability, as correlates of the task and social-emotional dimensions (see Table 1). The other two are consequences: *outcomes* are the behavioral and material results, such as selection, promotion, compensation, or dismissal, and *adaptation* refers to the adjustments a target makes in response to role incongruity, such as self-regulation or role reframing.

Table 1

Dimensions and Attributes Underlying Role Congruity Evaluations and Common Organizational Correlates

Dimension	Subdimension	Operationalizations	Correlates / Outcomes
Task	Competence/ability (task capability)	Intelligent, capable, skilled, knowledgeable, efficient, expert	Task performance, credibility, leadership effectiveness, trust
	Assertiveness/dominance (goal pursuit)	Confident, ambitious, determined, influential, takes initiative, competitive	Leadership emergence, perceived status, initiative-taking
Social-emotional	Warmth/friendliness (social approachability)	Friendly, sociable, likable, empathetic, team-oriented	Interpersonal attraction, social preference, relational trust
	Morality/integrity (intentions and ethics)	Honest, trustworthy, fair, ethical, sincere, loyal, authentic	Moral trust, inclusion, reputation, justice perceptions

Note. Dimensions and subdimensions following Abele et al. (2016, 2021). The task and social-emotional labels follow the classic distinction in research on groups and leadership (Bales, 1950); Abele et al. (2021) refer to these higher-order dimensions as vertical and horizontal. We retain the subdimensions and report them separately; the two dimensions organize rather than collapse them. The final column lists illustrative correlates of each subdimension (Abele & Wojciszke, 2014).

Role (In)congruity of Artificial Agents

Role congruity theory has been expanded beyond its original focus on gender and leadership to encompass a broader range of social roles and characteristics (for a review, see del Carmen Triana et al., 2024). For instance, role incongruity can emerge when organizational role expectations conflict with attributes associated with race (e.g., Paluch & Shum, 2024; Sy et al.,

2010), age (Daldrop et al., 2025; Diekman & Hirnisey, 2007), or sexual orientation (Liberman & Golom, 2015; Rule et al., 2016).

We extend this theory to explain how people judge *artificial agents* in organizational roles. While role congruity theory traditionally focuses on human social categories, artificial agents have become salient social actors in organizations (Anthony et al., 2023; Harris-Watson & Larson, 2026). This continues the "computers are social actors" (CASA) paradigm, which showed that people apply human social rules and expectations to computers and other artificial agents (Nass et al., 1994, 1996). People hold mental models about robotic and algorithmic agents (Banks, 2020; Cena & Grasso, 2026) and apply social perceptions to them (Nass & Moon, 2000). Because they evaluate these agents along the same dimensions that guide human judgments, such as competence and warmth (e.g., Harris-Watson et al., 2023; McKee et al., 2024), artificial agents can be perceived as congruent or incongruent with their organizational roles, much as human role occupants are.

Artificial agents increasingly occupy a range of organizational roles, including follower, partner, and even leader roles (e.g., Tsai, Marshall, et al., 2022; Wesche & Sonderegger, 2019). Whether an agent is perceived as congruent or incongruent therefore depends on what its role prescribes, which we turn to next.

Organizational Roles: Leader, Follower, and Partner

Organizational roles function as shared cognitive schemas that define expectations for behavior and abilities within hierarchical and collaborative structures (Anglin et al., 2022). Research on implicit leadership and followership theories demonstrates that people hold mental models of leaders and followers: sets of attributes and behaviors that describe ideal members of each category (Lord et al., 1984, 2020).

Leaders are expected to provide direction, coordination, and influence and to score high on both the task and the social-emotional dimension (Junker & van Dick, 2014). On the task dimension, leaders should be competent, able to analyze, decide, and execute effectively, and appropriately assertive; assertiveness is seen as ideal when it guides and coordinates action but counter-ideal when expressed as dominance (Benson et al., 2024; Offermann & Coats, 2018). On the social-emotional dimension, leaders are expected to show moral integrity, that is, fairness, honesty, and responsibility for others' welfare, alongside warmth (Den Hartog et al., 1999; Offermann & Coats, 2018). These social-emotional qualities are central to ideal-leader expectations and can even take precedence over assertiveness in followers' preferences (Daldrop et al., 2025; Ponce de Leon & Bailey, 2025).

Followers are expected to be dependable, conscientious, and loyal, carrying out directives and supporting collective goals (Junker et al., 2016; Sy, 2010). On the task dimension, expectations are modest: followers need enough competence to perform reliably, while assertiveness is expected to stay low, since dominance or overt self-promotion contradicts the deferential nature of followership (Bastardoz & Van Vugt, 2019; Benson et al., 2024). On the social-emotional dimension, the central expectations are warmth and supportive engagement, centered on cooperation rather than the moral authority expected of leaders (Benson et al., 2024; Junker et al., 2016).

People also hold expectations of partners, including teammates and colleagues, who collaborate as equals rather than within a hierarchy of authority. On the task dimension, expectations are moderate: partners should be competent enough to contribute reliably and to coordinate and communicate with others, but without the directive authority expected of leaders (Stevens & Campion, 1994). On the social-emotional dimension, warmth carries particular

weight among equals rather than the deference expected of followers; people favor warm, cooperative colleagues and, when choosing whom to work with, interpersonal liking can even outweigh competence (Casciaro & Lobo, 2008).

We have outlined what observers expect of leaders, followers, and partners. But how do people assess the fit between an artificial agent and the expectations of these roles? We argue that observers rely on the artificial agent's design-based social cues to ascribe attributes that make it appear congruent or incongruent with leader, follower, or partner roles.

From Design-Based Cues to Ascribed Attributes

Design-based social cues are the appearance and behavioral features designed into an artificial agent that signal its social presence to observers (Blut et al., 2021; Liew & Tan, 2021). These cues fall into two broad categories: appearance and verbal and non-verbal behavior. Appearance cues capture embodiment, an agent's form, degree of human likeness, and movement quality, ranging from disembodied algorithmic systems and voice-based assistants to virtual avatars and physically embodied robots (Glikson & Woolley, 2020; Liew & Tan, 2021). *Verbal* and *non-verbal behavioral cues* include communicative features such as voice, language style, responsiveness and turn-taking, gaze, gesture, facial expression, and labeling (Blut et al., 2021; Liew & Tan, 2021). Such communicative cues, for example polite or empathic phrasing, increase perceived social presence and human likeness (Banks et al., 2023; Jeon, 2022).

Design-based social cues, particularly those conveying social presence and human likeness, may enable people to infer specific attributes of the agent (Epley et al., 2007; Liew & Tan, 2021). One potential route from social cues to ascribed attributes runs through mind perception: observers infer the extent of the agent's mind along two capacities, agency (i.e., the capacity for self-control and planning) and experience (i.e., the capacity to feel and sense; Gray

et al., 2007; Waytz et al., 2010). These two capacities map loosely onto the same task and social-emotional dimensions that structure role expectations (e.g., Au & Ng, 2021). On the task dimension, perceived agency corresponds most clearly with competence and assertiveness. On the social-emotional dimension of role attributions, perceived experience corresponds with warmth, whereas both perceived agency and perceived experience map onto morality, the second facet of the social-emotional dimension (Gray et al., 2012). For example, social cues that signal feeling, such as a humanlike voice or empathic language, raise perceived experience as part of an artificial agent's mind perception and, in turn, its warmth and benevolence attributions (Koban & Banks, 2024; Yam et al., 2021). Table 2 summarizes how specific design features map onto these mind perceptions and ascribed attributes.

Recent research increasingly uses mind perception to explain how people relate to artificial agents at work: as a framework for human-agent teams (Harris-Watson & Larson, 2026), as a driver of how employees craft their relationships with industrial robots (Goštautaitė et al., 2026), and as an account of threat and fear of replacement across professional roles, such as doctors and judges, and personal ones, such as companions (Dong et al., 2026; Yam et al., 2025). Part of this work already draws on a lack-of-fit logic: whether people accept or resist a machine in a given role depends on how well the mind they ascribe to it matches the mind the role is seen to require (Dong et al., 2026; Yam et al., 2025).

Our review applies a role-congruity lens beyond mind perception, to role (in)congruity more broadly, comparing what a role prescribes with the attributes observers infer from an artificial agent's design-based cues. This conceptual background motivates the overarching research question that guides our integrative review: How do role-based expectations and design-

based inferences fit together for artificial agents in organizational leader, follower, and partner roles?

Table 2

Examples of Design-Based Cues, Mind Perception, and Ascribed Attributes

Design feature	Salient cues	Mind perception	Ascribed attributes
Appearance (embodiment)			
Embodiment	Humanoid form, motion fluency	Agency increases; experience may increase when expressivity is naturalistic	Competence and assertiveness; potentially warmth with rich expressivity
Verbal and non-verbal behavior			
Voice and language	Natural prosody, empathic wording, first-person phrasing	Experience increases (felt intent, care)	Warmth and morality
Responsiveness and gaze	Turn-taking, gaze, repair behavior, low latency	Agency increases and social presence increases	Competence and warmth
Labeling or name	"AI-generated," "approved by leader," "assistant" framing	Shifts perceived responsibility and intent	Can increase competence or decrease warmth depending on authenticity and role signals

Note. The table provides a conceptual synthesis of prior work: design features and salient cues summarize meta-analytic and review findings on embodiment and communicative behavior (Blut et al., 2021; Liew & Tan, 2021), linked to mind-perception theory (Gray et al., 2007; Waytz et al., 2010) and the task/social-emotional evaluation dimensions (Abele et al., 2016, 2021). This synthesis informed the coding scheme of our review (see Method).

Method

To answer our research question, we conducted a systematic review of the empirical literature on how people evaluate artificial agents in organizational roles. The Online Supplement, coding materials, and complete coded dataset are available on the Open Science Framework (https://osf.io/582jn/ overview).

Literature Search

We searched PsycINFO, Business Source Premier, and IEEE Xplore in June 2025, combining artificial-agent terms with social-perception constructs, and supplemented the database results with backward and forward citation tracing and targeted journal searches (full search string and per-source counts in Section A of the Online Supplement).

Inclusion and Exclusion Criteria

We included peer-reviewed empirical studies with human participants that examined an artificial agent (i.e., robot, AI system, avatar, or algorithm) in an organizational or task context and reported at least one perception of, or evaluative or acceptance response to, the agent in a leader, follower, or partner role. We restricted the review to studies that frame artificial agents as occupants of organizational roles, where role prescriptions are at stake. We excluded purely technical or conceptual papers, consumer-only studies without a work or team setting, and studies whose measured outcomes concerned only the human worker, third parties, or a non-agent object (full criteria in Section A of the Online Supplement).

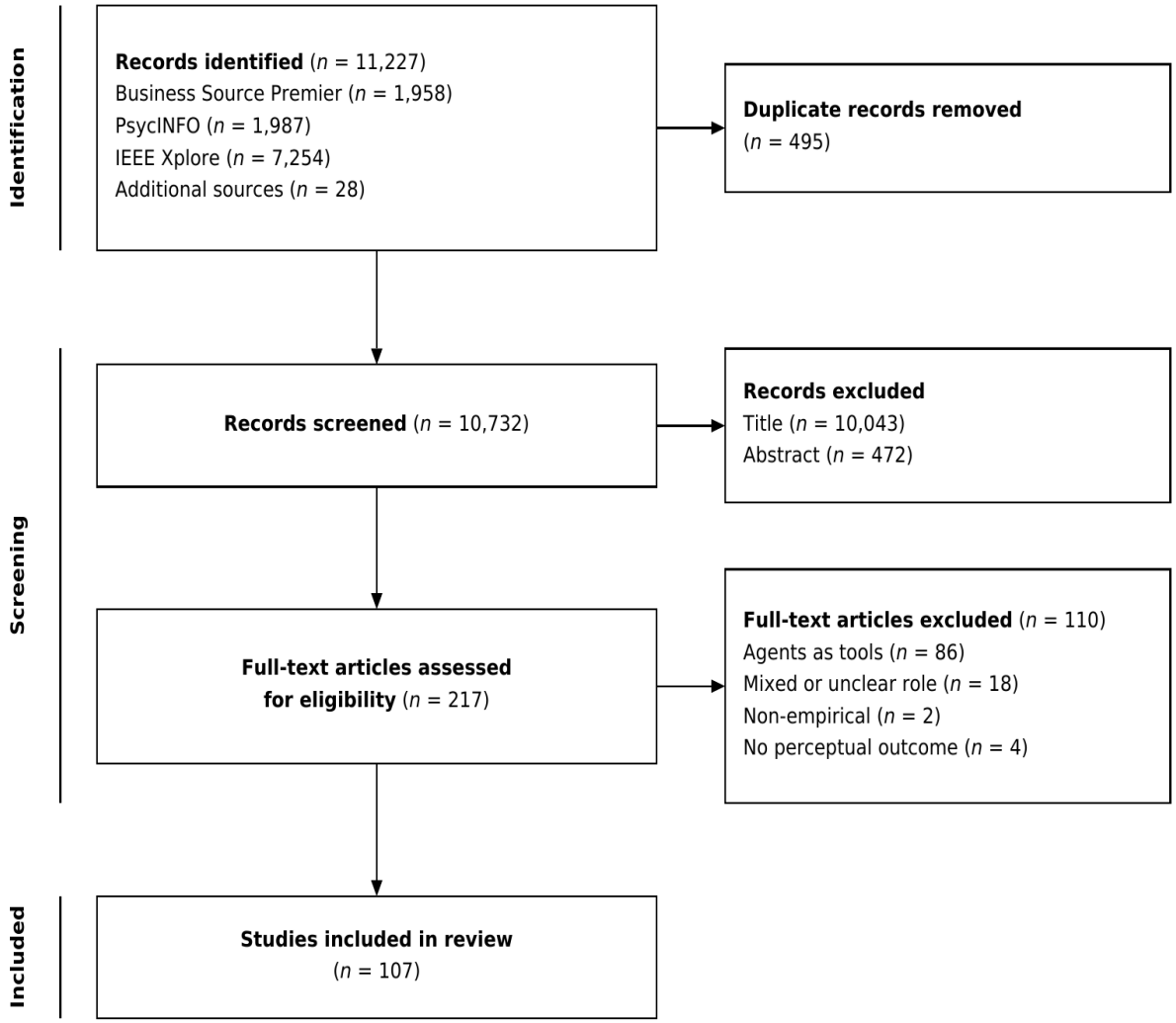
Screening and Selection

Screening followed PRISMA guidelines (Page et al., 2021): after deduplication, records were screened at title, abstract, and full text against these criteria, manually, with uncertain cases resolved by the author team and no automated tools used to exclude records. Studies excluded at

full-text screening are summarized in Section E of the Online Supplement. Figure 1 shows the resulting flow.

Figure 1

PRISMA Flow Diagram of the Identification and Screening Process



Note. Figure 1 is presented in accordance with PRISMA guidelines and details the flow of records through identification, screening, eligibility, and final inclusion (Page et al., 2021).

Coding Procedure and Scheme Development

Consistent with the iterative nature of integrative reviews (Cronin & George, 2023), a preliminary coding scheme developed by two trained research assistants on overlapping subsets was revised and applied to all included studies by the lead author. For each study we coded the agent's assigned role, that is, leader, follower, partner, or team member (Tsai, Marshall, et al., 2022); its design-based cues, both appearance and verbal and non-verbal behavior (Blut et al., 2021; Glikson & Woolley, 2020); the evaluation on the task (competence, assertiveness) and social-emotional (warmth, morality) dimensions (Abele et al., 2021); responses to role (in)congruity; and contextual variables (domain, research design, and sample region). The full coding documentation, including the codebook and scheme development, is provided in the Online Supplement (Sections B and C), and descriptive characteristics of the included studies are reported in Section D.

Results

We reviewed 107 empirical studies to identify evaluations indicating role congruity or role incongruity. Guided by role congruity theory (Eagly & Karau, 2002), we identified role incongruity when evaluations are inconsistent with the attributes a role is expected to require. We classified evaluations as incongruent when studies report lower evaluations on role-central facets (e.g., lower competence or morality for leaders; lower reliability-linked competence for followers; lower warmth or morality for partners) or evaluations that signal a violation of role expectations (e.g., excessive assertiveness for followers). In contrast, we classified evaluations as congruent when they indicate alignment with role-central facets and do not show evidence of role violations.

Using this rule set, we examine how an agent's assigned role and its design-based social cues relate to evaluations, and we summarize recurring patterns of congruity and incongruity by role. Summary tables present representative studies for agents as leaders (Table A1), followers (Table A2), and partners or teammates (Table A3); the full coded dataset is available on the Open Science Framework (https://osf.io/582jn/).

Overview and Descriptive Analysis

The review reveals a recent acceleration in research on perceptions of artificial agents in organizational or task-related roles, with 89% of studies published in 2020 or later (95 of 107). Work prior to 2015 primarily consisted of social robotics and trust experiments (e.g., Salem et al., 2013), while the period from 2016 to 2019 marked the first integration of workplace-relevant contexts linking human likeness to evaluations of competence and warmth (Castro-González et al., 2016; Lee, 2018). Research starting in 2020 increasingly addressed algorithmic and robotic leadership (e.g., Jago et al., 2024; McGuire & De Cremer, 2023), teamwork and collaboration (e.g., Fraune et al., 2020; Harris-Watson et al., 2023), and service or caregiving settings (e.g., Paluch et al., 2022; Yam et al., 2021). Across disciplines, leader-role research is concentrated in management and psychology, partner-and teammate-role research in robotics, engineering, and computer science, and follower-role research is spread more evenly across fields. Detailed distributions by publication year, disciplinary field, primary context, agent design type, research design, and sample region are provided in the Online Supplement Section D (Tables S5–S11).

Role Incongruity of Artificial Agents as Leaders

A recurring pattern emerges for artificial agents in leader roles: they are often credited with task competence but fall short on the warmth, benevolence, and legitimacy that leadership also requires. On the task dimension, artificial leaders are frequently judged competent. An AI avatar

delivering a vision was rated more competent, though less warm, than a human delivering the same vision (Buss et al., 2025); robotic managers are accepted more readily on performance-related than on relational aspects (Wolf & Stock-Homburg, 2026). Automated leaders can also be seen as comparable in ability and higher in consistency and transparency than human leaders (Höddinghaus et al., 2021).

The consistent deficit is on the social-emotional dimension. Artificial leaders and their decisions are perceived as less fair, trustworthy, and benevolent than human counterparts (Li & Bitterly, 2024; McGuire & De Cremer, 2023; Wesche et al., 2024), particularly for decisions seen to require human judgment (Lee, 2018). Anthropomorphic and agentic design cues do not reliably close this gap and can even widen it: humanlike robot supervisors elicit greater perceived agency and, in turn, perceptions of abuse and supervisor-directed retaliation when delivering negative feedback (Yam et al., 2022), and employees comply less with unethical instructions from AI than from human supervisors (Lanz et al., 2024).

Overall, the evidence points to a competent-but-cold pattern: artificial leaders can meet competence expectations yet fall short on the benevolence, fairness, and legitimacy that confer standing in leader roles, an incongruity concentrated on the social-emotional dimension. We present sample studies for agents as leaders in Table A1.

Role Congruity of Artificial Agents as Followers

For artificial agents in follower or subordinate roles (e.g., service, assistive, or support functions), the evidence points to congruity between design-based cues and follower-role expectations. When design features signal social responsiveness without asserting authority, agents are generally evaluated as both warm and competent (Yang et al., 2025). Anthropomorphic cues that convey emotional capacity raise perceived experience and

satisfaction (Yam et al., 2021) and foster relational investment, including parasocial bonds with the agent (Noor et al., 2022), while warm, attentive service agents enhance psychological ownership and revisit intentions (Ruiz-Equihua et al., 2023). Functional cues that signal reliability and capability, in turn, build the competence-based trust that drives adoption (Kraus et al., 2024).

The pattern reverses mainly when human-likeness becomes too pronounced: a highly anthropomorphic, physically present driving agent lowered perceived control and cognitive trust relative to a virtual one (Cheng et al., 2022). On balance, the follower evidence supports role congruity, with most design-based cues aligning with follower-role expectations for warmth, responsiveness, and dependable performance. This evidence base is concentrated in service and caregiving contexts (23 of 37 follower studies, 62%; 18 retail or service and 5 healthcare) and includes only two internal-office studies, so we see the congruity pattern as most directly applicable to service and caregiving settings. We present sample studies for agents as followers in Table A2.

Role (In)congruity of Artificial Agents as Partners and Teammates

For artificial agents positioned as partners or team members, evaluations are mixed and depend on boundary conditions. Agents that signal prosocial orientation and competence are generally accepted: prosocial (high social-value-orientation) agents are rated warmer, and warmth drives partner preference and continued cooperation beyond objective performance (McKee et al., 2024). A more humanlike agent framed as a teammate also elicits greater initial trust and willingness to collaborate (Mirowska & Arsenyan, 2025).

Acceptance erodes when an agent falls short of the expectations its framing creates. A “colleague” framing raises expectations that an error-prone agent fails to meet, creating

frustration (Einola et al., 2024); moral violations reduce trust more sharply than equally consequential performance failures (Rezaei Khavas et al., 2024); and strongly anthropomorphic collaborative robots raise perceived intelligence and, in turn, perceived threat and negative attitudes (Liao et al., 2024). Team composition matters as well: interpersonal trust and identification are lower in two-member human-AI dyads than in all-human dyads, though this pattern was attenuated in three-member teams (Georganta & Ulfert, 2024).

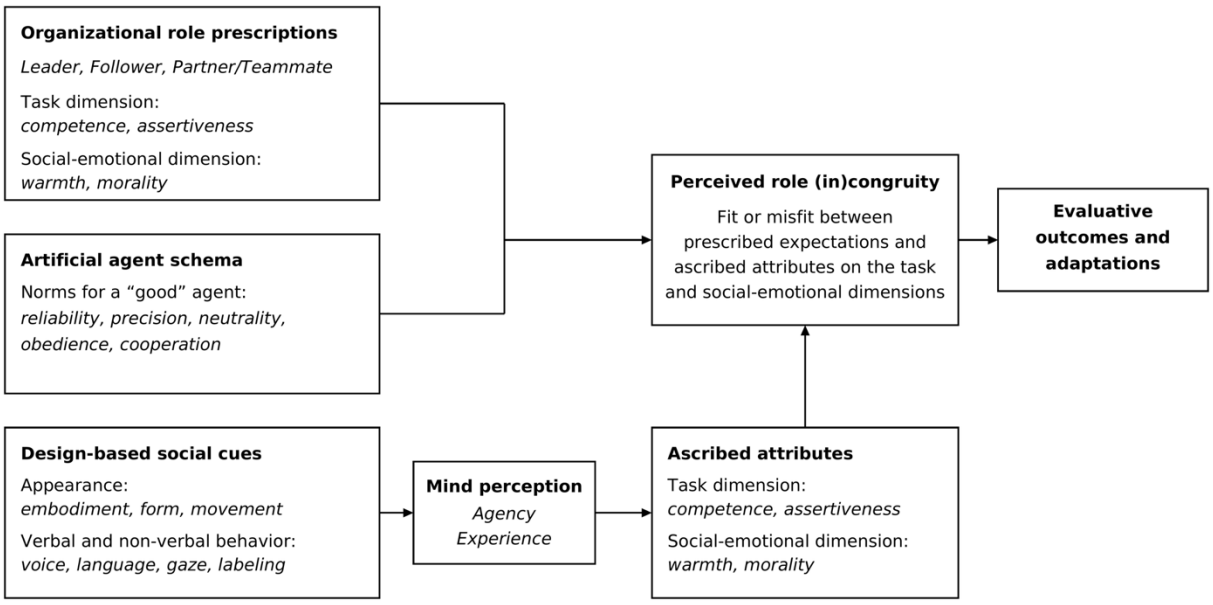
Collectively, these findings tie evaluations of partner and teammate agents to the interaction between role prescriptions for reciprocity and cooperation and the attributes inferred from design cues. Role congruity emerges when cues signal enough social presence and experience to support trust and coordination without so much agency or autonomy that they violate relational expectations; role incongruity emerges when anthropomorphic or agentic cues exceed those expectations, eliciting discomfort or threat. We present sample studies for agents as partners or team members in Table A3.

Toward the Artificial Agent Role Congruity (AARC) Model

We develop the Artificial Agent Role Congruity (AARC) model by extending role congruity theory (Eagly & Karau, 2002) to artificial agents. AARC formalizes how observers compare prescribed expectations of organizational roles with ascribed attributes inferred from an artificial agent's design (Figure 2), producing perceptions of role congruity or role incongruity. We present AARC as a conceptual model (Rousseau, 2026) that invites empirical testing and refinement. The model organizes the evidence reviewed above and identifies when, how, and why artificial agents are evaluated differently across organizational roles.

Figure 2

Artificial Agent Role Congruity (AARC) Model of Role-Based Expectations and Design-Based Social Cues



Note. The AARC model illustrates how role-based expectations and design-based social cues shape evaluations of artificial agents. Two sets of expectations, the prescriptions of the organizational role and the artificial-agent schema, are compared with the attributes ascribed to the agent. Design-based social cues are read, through mind perception (agency and experience), as ascribed attributes along two dimensions of social perception. Perceived role (in)congruity is the fit or misfit between these ascribed attributes and what the role and schema prescribe, which in turn shapes evaluative outcomes and adaptations such as acceptance, trust, and legitimacy.

As our review through a role incongruity lens suggests, the combination of roles and design features determines how artificial agents are evaluated. Whereas role congruity is attained when role and design match, two types of role incongruity are conceivable. In roles > design configurations, the expectations attached to functional or organizational roles exceed what the agent's design appears to offer. Agents framed as leaders, colleagues, or partners are expected to display empathy and moral intent, yet their design typically signals limited social presence or experience. This shortcoming may generate a deficit on the social-emotional dimension, leading

observers to perceive them as tools or morally inadequate actors (Einola & Khoreva, 2023; McGuire & De Cremer, 2023). In contrast, design > roles configurations occur when an agent's design implies more agency, dominance, or human likeness than its role requires. For example, anthropomorphic or socially assertive cobots may appear overly competent or intrusive in subordinate or tool roles, causing threat or unease (Liao et al., 2023, 2024; Stanton & Stevens, 2017). Both patterns reduce congruity between role-based prescriptions and design-based social cues, thereby defining the core antecedent of role incongruity in line with the AARC model.

Taken together, the AARC model adds three elements beyond role congruity theory applied to a new class of role occupants. First, in artificial agent evaluation design-based social cues replace social-category cues as the source of ascribed attributes. What gender, age, or social class signal for role congruity evaluations in research on human role occupants, embodiment, voice, language, and labeling signal for artificial agents; because designers can control these cues, ascribed attributes become a design parameter rather than a demographic given. Second, mind perception bridges cues and evaluation: perceived agency feeds the task dimension and perceived experience the social-emotional dimension, so the model predicts that experience cues carry more evaluative weight in roles whose prescriptions emphasize warmth and morality. Third, the artificial agent schema—that is, the shared expectations about how robots or algorithms should act and relate (e.g., reliability, precision, obedience, neutrality, cooperation; Banks, 2020)—constitutes a second layer of (in)congruity, so an agent can satisfy its organizational role prescriptions yet violate this artificial agent schema. Together, these elements specify the extent of perceived misfit and why it emerges, rather than merely noting that artificial agents differ from human role occupants.

Discussion

With this review, we asked how role-based expectations and design-based inferences jointly shape evaluations of artificial agents in organizational contexts. Three patterns emerged. First, artificial agents in leader roles often face role incongruity, particularly on the social-emotional dimension. Second, agents in follower roles generally elicit congruity, especially in service and caregiving settings. Third, congruity in partner roles depends on design and context. These patterns represent the AARC model's mechanism: role prescriptions and the attributes inferred from design jointly shape evaluations on the task (competence, assertiveness) and social-emotional (warmth, morality) dimensions (Abele et al., 2021). Across roles, incongruity took two forms: either role demands exceeded what the design conveyed (roles > design), or the design implied more agency or human likeness than the role required (design > roles). Notably, humanlike design rarely repairs the leader-role deficit; if anything, greater human likeness intensifies discomfort by amplifying perceived agency without a matching gain in perceived moral understanding, consistent with accounts of AI trust as resting jointly on morality and performance (Everett et al., 2026).

Theoretical Implications

Our review suggests three theoretical extensions. The first links the model to the mind-perception account of machine replacement. That literature frames reactions to machines as a matter of mind–role fit, the correspondence between the mind ascribed to an agent and the mind a role is taken to require (Dong et al., 2026; Yam et al., 2025). In role-congruity terms, this is a lack of fit between prescribed and ascribed attributes in which the ascribed attributes are the agent's perceived mind; mind–role (mis)fit is therefore a special case of role (in)congruity rather than a separate mechanism. Because perceived mind comprises two capacities, agency and

experience, the dimension on which the lack of fit appears is predictable. In leader roles, incongruity arises when design cues convey agency but little experience, so that agents are judged competent yet lacking in warmth and moral intent (e.g., McGuire & De Cremer, 2023; Rubin et al., 2025). In follower roles, the lack of fit reverses, as mind-like cues exceed what a controllable, impersonal role prescribes (e.g., Appel et al., 2020). In partner and teammate roles, a comparable surplus of agency or deficit of experience conflicts with the reciprocity and warmth these roles prescribe (e.g., Banks et al., 2023). Framed this way, the mind-perception account of machine replacement is subsumed within the broader role-congruity logic of the AARC model.

Within our AARC model, an agent's design not only constitutes a set of cues but can also serve as a baseline "artificial agent role" schema. When an agent's behavior or perceived attributes violate this artificial agent schema, observers may perceive a role violation that is comparable to violating a social role such as the gender role in role congruity theory, where prejudice can arise from failing to meet the prescriptions of a social role over and above the functional role itself (Eagly & Karau, 2002). Consequently, negative reactions may stem not only from failing to meet leader-, partner-, or follower-role prescriptions but also from not fulfilling the norms of being a "good robot" or "trustworthy algorithm." Empirical evidence illustrates this dynamic: agents that make mistakes (Dietvorst et al., 2015; Weidemann & Russwinkel, 2021), act unpredictably (Salem et al., 2013), cheat (Herath et al., 2018), or display humanlike affect or behavior such as anger (Herdel et al., 2024), dialect (Kühne et al., 2023), or humming (Liu et al., 2023) violate this artificial agent schema. Such violations can lead observers to downgrade agents to malfunctioning tools or react with abjection (Banks et al., 2023; Einola & Khoreva, 2023), while in other cases they "upgrade" agents to social actors due to an increase in mind perception (as this behavior is more humanlike). Integrating this artificial agent schema into the

AARC model adds a second dimension of role incongruity in artificial agent evaluation, arising not only from mismatched organizational roles but also from deviations from the normative expectations embedded in the artificial agent role itself.

The reviewed evidence also points to a contingency that the model does not yet formalize: the context in which a role is embedded shapes how strongly its prescriptions emphasize the task versus the social-emotional dimension. The composition of our database illustrates this. For agents in follower roles, evidence is concentrated in service and caregiving settings (23 of 37 studies, 62%), where role prescriptions emphasize warmth and responsiveness; the congruity we observe is therefore most directly a statement about service contexts. Leader-role evidence comes predominantly from internal employee and management settings (10 of 17 studies, 59%), where moral and relational prescriptions are most demanding, which sharpens the incongruity we observe. We propose that artificial agents will be evaluated as less incongruent in leader roles whose context emphasizes task prescriptions over social-emotional ones, such as data-intensive allocation or monitoring tasks, and as more incongruent in follower or partner roles whose context elevates social-emotional prescriptions, such as care work. We therefore treat the generalizability of our role-specific conclusions accordingly (see the scope and boundary conditions in the Online Supplement, Section A).

Practical Implications

Our findings translate into concrete guidance for managers deploying artificial agents and for designers building them (Chhillar & Aguilera, 2022; Kellogg et al., 2020). For managers, three priorities stand out. First, retain human moral oversight in algorithmic management: hybrid human-in-the-loop arrangements increase trust and perceived fairness compared to full automation (De Cremer & McGuire, 2022; Kares et al., 2023). Second, configure leadership as

augmentation rather than replacement: combining algorithmic capabilities with human judgment addresses the moral and relational deficits that design alone cannot repair (Raisch & Fomina, 2025; Raisch & Krakowski, 2021). Field evidence shows that the governance of such arrangements can be configured so that organizations and workers benefit jointly (Kellogg, 2022). Third, specify the role frame explicitly when introducing an agent: role ambiguity produces unmet "colleague" expectations and disappointment, whereas clear framing aligns observer expectations with what the agent can deliver (Einola et al., 2024).

For designers, three principles follow. First, do not assume that anthropomorphizing agents intended for leader roles will resolve perceived moral deficits: humanlike cues may amplify rather than reduce role incongruity (McGuire & De Cremer, 2023; Yam et al., 2022). Second, design for social presence rather than human likeness: responsiveness, relational language, and transparent identity framing build acceptance without uncanny-valley risks. Third, match design cues to the role: human likeness is well received in follower roles, accepted conditionally in partner roles, and rarely fitting in leader roles.

These principles rest on convergent evidence: social presence matters more for acceptance and cooperation than humanlike appearance (Elshan et al., 2022). Presence cues such as relational language, immediacy, and transparent identity framing raise trust and willingness to rely on the system. Humanlike appearance alone, however, may be insufficient to support social presence in human–AI collaboration (Siemon et al., 2026). Strong cues can also backfire: excessive perceived autonomy reduces cooperation through a reactance process of "agency-cued abjection," a feeling of unease and rejection (Banks et al., 2023), and reading an agent's outputs as signs of "reasoning" encourages false confidence in its capability (Kambhampati et al., 2026). More broadly, these prescriptions show what human-centric deployment of artificial agents

requires in practice: respecting the role expectations through which workers grant legitimacy to a technology, so that efficiency gains do not come at the cost of the warmth, morality, and fairness that make any role occupant acceptable.

Limitations

Our review has limitations. First, the database covers peer-reviewed journal articles, so cutting-edge findings from conference proceedings in cognate fields such as computer science and human-robot interaction may be missing. Future reviews could broaden the inclusion criteria to high-quality conference proceedings to mitigate this. Second, the final coding scheme was applied by the lead author after a two-phase scheme development. Although the procedure is described in the Method, the inter-coder checks are documented in the Online Supplement, and the full coding is openly available, single-coder application remains a constraint. Third, the evidence base is unevenly distributed across contexts: follower-role evidence concentrates in service and caregiving settings and leader-role evidence in internal organizational settings. This calls for further investigations into the generalizability of role-specific conclusions (see the scope and boundary conditions in the Online Supplement, Section A).

Opportunities for Future Research

Building on the AARC model and our review findings, we outline opportunities in three areas: theory development, measurement and design, and organizational consequences. Table 3 provides an overview of these opportunities, which we elaborate in the following subsections.

Table 3
Opportunities for Future Research on Artificial Agents in Organizational Roles

Topic	Core idea
Theory development	
Role congruity vs. role ambiguity or role conflict	Distinguish when responses to artificial agents reflect clear role (in)congruity evaluations versus *ambiguity* or *conflict* between competing role schemas.
Reversing human-machine hierarchies	Examine *status incongruity* as a pathway to resistance when artificial agents occupy high-status roles such as supervisor, evaluator, or manager.
Measurement and design	
Direct tests of role (in)congruity	Move from indirect inference to designs that measure both role prescriptions and agent attributes and test role incongruity causally.
Research design effects	Treat research design as a boundary condition for evaluations of artificial agents and compare scenario-based, case-based, and field designs.
Baseline comparisons and benchmarks	Introduce *neutral* and *treatment-as-usual* baselines to separate true role-design effects from artifacts of relative contrasts.
Organizational consequences	
Team composition and in-group vs. out-group categorization	Treat *in-group* vs. *out-group* categorization as a boundary condition for role (in)congruity in human-agent teams.
Recognition-based vs. inference-based processes	Explore when performance evidence can shift perceptions from *recognition-based* to *inference-based judgments* of artificial leaders.
Spillover to third parties	Measure how artificial agents change third-party evaluations of the humans who work with or under them.
Automation vs. augmentation	Use the *automation-augmentation* tension to rethink where artificial agents should lead, support, or be constrained by human oversight.

Theory Development

Role Congruity Versus Role Ambiguity or Conflict. Role theory distinguishes a *structural-functional view*, in which roles are relatively fixed positions with prescriptive expectations, and a *symbolic-interactionist view*, in which roles are negotiated in interactions (Anglin et al., 2022). In the reviewed research, evaluations of artificial agents in leader roles typically follow the structural-functional logic of role congruity theory (del Carmen Triana et al.,

2024; Eagly & Karau, 2002): Agents are formally introduced as supervisors or managers, and observers compare prescribed leader expectations with attributes ascribed from design-based cues to judge role congruity or role incongruity. In contrast, studies of artificial agents in partner and follower roles often report patterns that we interpret as *role ambiguity* or *role conflict*. We derive this from evidence that participants waver between different role labels and expectations, or explicitly question what the agent "really is," for example when an algorithm formally introduced as a "colleague" was later described by employees as a "shiny tool" when partner role expectations were not met (Einola et al., 2024). Likewise, platform algorithms that fulfill tool-like functions (e.g., navigation) and supervisory functions (e.g., giving directives and enforcing performance ratings) blur tool, follower, and leader categories and evoke competing role expectations (Tarafdar et al., 2023). We encourage future research to clarify when evaluations are primarily driven by structural-functional role prescriptions and when they are shaped by negotiated role meanings, and to specify boundary conditions under which role incongruity, role ambiguity, or role conflict best captures responses to artificial agents. Such clarification matters because role prescriptions themselves may shift: some argue that artificial agents could come to address followers' basic psychological needs better than human leaders (Van Quaquebeke & Gerpott, 2023), which would make the leader-role deficits we document contingent rather than constant.

Reversing Human–Machine Hierarchies. While role (in)congruity represents the primary evaluative mechanism in the AARC model, alternative pathways may also explain resistance toward artificial agents. One such pathway is status incongruity, which occurs when an agent's assigned hierarchical position conflicts with established social norms and expectations (Rudman et al., 2012). Leader roles imply higher status and influence, whereas artificial agents are

typically viewed as technologies that support or accompany human workers (Lee & See, 2004). When this hierarchy is reversed, such as when an artificial agent leads or evaluates human employees, observers may perceive the hierarchy as inappropriate and question the agent's legitimacy (e.g., Jeon, 2022). Similarly, algorithmic management can reduce perceptions of job complexity and status of human organizational members, reinforcing this incongruity between human subordinates and machine supervisors (e.g., Jago et al., 2024). Future research could examine how mind-role and status incongruities jointly influence acceptance across hierarchical contexts.

Measurement and Design

Direct Tests of Role (In)congruity. There are still almost no direct tests of role (in)congruity or systematic baselines for specific agent designs, with only a few notable exceptions. Some recent studies begin to operationalize role (in)congruity or mind-role (mis)fit by measuring both role requirements and perceived agent characteristics, showing that mismatches between the two can shape perceptions and reactions to artificial agents (Dong et al., 2026). Further, evidence establishing causality for role incongruity as the underlying mechanism remains limited; one exception is experimental work that manipulates mind perception, agency and experience, to test their direct impact on behavior and moral judgment (Lanz et al., 2024). These early efforts are promising but underscore the need for more empirically rigorous evidence linking mind–role fit to target evaluations and outcomes. In addition, methods such as polynomial regression and response-surface analysis (Tsai, Kim, et al., 2022) could provide a formal approach to testing AARC's proposed mechanisms by explicitly measuring how both role prescriptions and design cues jointly predict perceived fit, enabling direct tests of the "roles > design" and "design > roles" patterns identified in this review.

Research Design Effects. Findings on artificial agents may vary with research design. Organizational case studies reveal mixed reactions, with negative evaluations of artificial agents coexisting with optimism about their future capabilities or positive implications for employees' work meaningfulness (e.g., Einola et al., 2024; Goštautaitė et al., 2024). Scenario-based vignette studies may yield more negative responses, potentially because participants imagine rather than experience the interaction (e.g., Liao et al., 2024; Savela et al., 2021), whereas field experiments that combine real task engagement with experimental control suggest that human–agent differences may diminish under real behavioral conditions (Dong et al., 2024). Research design may therefore function as a boundary condition: direct and sustained interaction may allow initial evaluations to become more differentiated as people gain experience with an artificial agent. Future work should move beyond short vignettes, "Wizard of Oz" experiments (where a hidden human controls the system), and other brief simulations toward designs built on real rather than hypothetical interaction. Even simple, well-controlled lab studies with genuine agent interaction would help, and longitudinal and field designs could track how trust, legitimacy, and cooperation develop over time in real workplaces.

Baseline Comparisons as Benchmark. Most research on artificial agents relies on relative comparisons such as human versus agent leaders or high versus low anthropomorphism, without including a baseline in which participants perform the task without guidance from either a human or an artificial agent. Although informative, such designs make it hard to see whether differences reflect genuine role–design effects or artifacts of the comparison. Evidence from our review underscores this concern. In experiments that included a self-directed or "passive" baseline, both human and robot guidance reduced perceived control relative to freely chosen actions, indicating that the loss of agency stemmed from being instructed rather than from the agent's identity

(Barlas, 2019). These findings align with recent methodological critiques that comparisons between two active conditions can produce pseudo effects that may obscure underlying mechanisms (e.g., Stark et al., 2026). Future research should therefore additionally incorporate no-treatment baselines, where app*ropriate*, alongside comparisons with a human leader performing the same task.

Organizational Consequences

Team Composition, In-Group Versus Out-Group Categorization, and Role Congruity.

Recent reviews of human–agent teams show that mixed teams frequently cause *in-group versus out-group categorization* processes, with artificial agents often treated as out-group members (Bezrukova et al., 2023; Wolf & Stock-Homburg, 2023). Examples from our review mirror this pattern: introducing robot teammates reduces identification and psychological attachment, particularly as the number of robot teammates increases (Savela et al., 2021), whereas collaboration improves when at least one additional human partner is present (e.g., Georganta & Ulfert, 2024; Musick et al., 2021). These out-group effects are conceptually distinct from, but consequential for, role (in)congruity of artificial agents as teammates. We therefore propose treating in-group versus out-group categorization as a boundary condition in our AARC model: out-group assignment may lower warmth ascriptions and moral expectations, or shift thresholds for judging fit, potentially overshadowing or amplifying role (in)congruity evaluations. Future research should jointly assess categorization and role-fit processes to clarify when responses to artificial teammates reflect out-group bias, role incongruity, or their combined influence.

Recognition-Based Versus Inference-Based Processes. Role incongruity of artificial agents in leader roles stems from recognition-based processes (e.g., Lord et al., 2020), where observers rely on design-based cues about who appears to “fit” a leader role. Yet, potential

perceptual biases raise a critical question: can demonstrated performance override these (possibly biased) perceptions? Early evidence suggests that algorithmic and AI-assisted managers can, in some contexts, perform effectively enough to challenge the biased judgments. For example, field studies show that algorithmic supervision can improve service quality and customer satisfaction, even if such gains come with trade-offs in efficiency (Gu et al., 2025). Large-scale experiments on human–AI collaboration similarly indicate that mixed human– AI teams achieve higher productivity and, in some domains, produce superior output quality relative to human-only teams (Ju & Aral, 2025). These patterns imply that evidence about an agent's actual performance can shift social perception from a recognition-based process, which relies on whether the agent "looks like" a leader, to an inference-based process that places more weight on what the agent demonstrably achieves (Lord et al., 2020). Future research could examine when and how performance signals can attenuate role incongruity, and under what conditions outcome-based credibility can compensate for role incongruity in establishing leader legitimacy for artificial agents (over time).

Spillover to Third Parties. Artificial agents may reshape how third parties perceive the humans who work with or under them. Early evidence suggests that being managed by an algorithm can change how employees are evaluated by others (Schweitzer & De Cremer, 2024). Recent evidence extends this pattern along the model's two dimensions: algorithmic management was associated with lower coworkers' affective (warm) trust in newcomers and higher cognitive (task-related) trust (Lin et al., 2026). Such spillover effects on coworkers, applicants, and external observers were rarely measured in the reviewed studies and represent an organizational-level consequence of role (in)congruity that merits systematic attention.

Automation Versus Augmentation. Our findings suggest that role incongruity for artificial agents acting as leaders is unlikely to be eliminated; it may instead reflect a deeper tension between automation and augmentation (Raisch & Krakowski, 2021). The more decision-making is automated, the greater the need for human oversight and interpretive leadership, so attempts to replace human leaders may increase the demand for distinctly human capabilities such as contextual judgment and social sensemaking. As we note in the practical implications, the remedy lies less in resolving incongruity through design or framing than in reconfiguring leadership into hybrid forms that combine algorithmic input with human judgment (Raisch & Fomina, 2025). Future research should examine how such hybrid architectures can balance analytical augmentation with human interpretive oversight to manage, rather than eliminate, the persistent role incongruity for nonhuman leaders.

Conclusion

Across organizational roles, artificial agents are evaluated very differently: those in leader roles face persistent incongruity, judged competent but short on the morality and legitimacy that leadership demands; those in follower roles show consistent congruity when their design signals responsiveness and reliability; and those in partner or teammate roles fall in between, with congruity conditional on context and design. Our Artificial Agent Role Congruity (AARC) model accounts for this variation: evaluations turn on the (in)congruity between role-based prescriptions and the attributes inferred from design-based cues along the task and social-emotional dimensions of social perception. By extending role congruity theory to nonhuman actors, AARC clarifies when artificial agents are judged as legitimate role occupants and where misfit biases evaluations and outcomes.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290–314. https://doi.org/10.1037/rev0000262

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness - Communion with warmth and morality. *Frontiers in Psychology*, *7*, 1810. https://doi.org/10.3389/fpsyg.2016.01810

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in Experimental Social Psychology*, 50, 195–255. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Anglin, A. H., Kincaid, P. A., Short, J. C., & Allen, D. G. (2022). Role theory perspectives: Past, present, and future applications of role theories in management research. *Journal of Management*, *48*(6), 1469–1502. https://doi.org/10.1177/01492063221081442

Anthony, C., Bechky, B. A., & Fayard, A.-L. (2023). "Collaborating" with AI: Taking a system view to explore the future of work. *Organization Science*, *34*(5), 1672–1694. https://doi.org/10.1287/orsc.2022.1651

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in Human Behavior*, *102*, 274–286. https://doi.org/10.1016/j.chb.2019.07.031

Asimov, I. (1950). *I, robot*. Gnome Press.

Au, R. H. Y., & Ng, G. T. T. (2021). Mind perception and stereotype attribution of corporations and charities. *British Journal of Social Psychology*, *60*(1), 271–293. https://doi.org/10.1111/bjso.12377

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Addison-Wesley.

Banks, J. (2020). Optimus primed: Media cultivation of robot mental models and social judgments. *Frontiers in Robotics and AI*, *7*, 62. https://doi.org/10.3389/frobt.2020.00062

Banks, J., Koban, K., & Haggadone, B. (2023). Avoiding the abject and seeking the script: Perceived mind, morality, and trust in a persuasive social robot. *ACM Transactions on Human-Robot Interaction*, *12*(3), 1–24. https://doi.org/10.1145/3572036

Barlas, Z. (2019). When robots tell you what to do: Sense of agency in human- and robot-guided actions. *Consciousness and Cognition*, *75*, 102819. https://doi.org/10.1016/j.concog.2019.102819

Bastardoz, N., & Van Vugt, M. (2019). The nature of followership: Evolutionary analysis and review. *The Leadership Quarterly*, *30*(1), 81–95. https://doi.org/10.1016/j.leaqua.2018.09.004

Benk, M., Kerstan, S., von Wangenheim, F., & Ferrario, A. (2025). Twenty-four years of empirical research on trust in AI: A bibliometric review of trends, overlooked issues, and future directions. *AI & Society*, *40*(4), 2083–2106. https://doi.org/10.1007/s00146-024-02059-y

Benson, A. J., Woodley, H. J. R., Jensen, L., & Hardy, J. (2024). Mapping the traits desired in followers and leaders onto fundamental dimensions of social evaluation. *Journal of Experimental Psychology: Applied*, *30*(4), 586–606. https://doi.org/10.1037/xap0000514

Berretta, S., Tausch, A., Ontrup, G., Gilles, B., Peifer, C., & Kluge, A. (2023). Defining human-AI teaming the human-centered way: A scoping review and network analysis. *Frontiers in Artificial Intelligence*, *6*, 1250725. https://doi.org/10.3389/frai.2023.1250725

Bezrukova, K., Griffith, T. L., Spell, C., Rice, V., & Yang, H. E. (2023). Artificial intelligence and groups: Effects of attitudes and discretion on collaboration. *Group & Organization Management*, *48*(2), 629–670. https://doi.org/10.1177/10596011231160574

Biddle, B. J. (1986). Recent developments in role theory. *Annual Review of Sociology*, *12*(1), 67–92. https://doi.org/10.1146/annurev.so.12.080186.000435

Bigman, Y. E., Briker, R., & Langer, M. (2026). Human–AI interaction research needs to be embedded in psychological theory. *Nature Reviews Psychology*, *5*(4), 233–234. https://doi.org/10.1038/s44159-026-00551-4

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632–658. https://doi.org/10.1007/s11747-020-00762-y

Boyle, M. (2025). *Meetings won't be the same when the boss sends an AI bot*. Bloomberg. https://archive.today/tpFdx

Buss, M., Metzger, M. J., & Kearney, E. (2025). Should leaders outsource visionary leadership to AI? The effects of AI vision delivery and perceived AI-generated vision content on follower perceptions and motivation. *Journal of Applied Social Psychology*, *55*(12), 1049–1062. https://doi.org/10.1111/jasp.70028

Casciaro, T., & Lobo, M. S. (2008). When competence is irrelevant: The role of interpersonal affect in task-related ties. *Administrative Science Quarterly*, *53*(4), 655–684. https://doi.org/10.2189/asqu.53.4.655

Castro-González, Á., Admoni, H., & Scassellati, B. (2016). Effects of form and motion on judgments of social robots′ animacy, likability, trustworthiness and unpleasantness. *International Journal of Human-Computer Studies*, *90*, 27–38. https://doi.org/10.1016/j.ijhcs.2016.02.004

Cena, F., & Grasso, F. (2026). Exploring users' mental models of conversational agents: A systematic review. *Behaviour & Information Technology*, 45(10), 2189–2212. https://doi.org/10.1080/0144929x.2025.2573436

Chamorro-Premuzic, T., & Ahmetoglu, G. (2016). The pros and cons of robot managers. *Harvard Business Review*, *12*, 2–5.

Cheng, P., Meng, F., Yao, J., & Wang, Y. (2022). Driving with agents: Investigating the influences of anthropomorphism level and physicality of agents on drivers' perceived control, trust, and driving performance. *Frontiers in Psychology*, *13*, 883417. https://doi.org/10.3389/fpsyg.2022.883417

Chhillar, D., & Aguilera, R. V. (2022). An eye for artificial intelligence: Insights into the governance of artificial intelligence and vision for future research. *Business & Society*, *61*(5), 1197–1241. https://doi.org/10.1177/00076503221080959

Cronin, M. A., & George, E. (2023). The why and how of the integrative review. *Organizational Research Methods*, *26*(1), 168–192. https://doi.org/10.1177/1094428120935507

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

Daldrop, C., Homan, A. C., & Buengeler, C. (2025). Too young to lead? Role incongruity explains age bias against young leaders. *The Leadership Quarterly*, *36*(4), 101878. https://doi.org/10.1016/j.leaqua.2025.101878

Dang, Q., & Li, G. (2026). Unveiling trust in AI: The interplay of antecedents, consequences, and cultural dynamics. *AI & Society*, *41*(1), 669–692. https://doi.org/10.1007/s00146-025-02477-6

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

del Carmen Triana, M., Song, R., Um, C. T., & Huang, L. (2024). Stereotypical perception in management: A review and expansion of role congruity theory. *Journal of Management*, *50*(1), 188–215. https://doi.org/10.1177/01492063231180836

Den Hartog, D. N., House, R. J., Hanges, P. J., Ruiz-Quintanilla, S. A., Dorfman, P. W., Abdalla, I. A., Adetoun, B. S., Aditya, R. N., Agourram, H., Akande, A., Akande, B. E., Akerblom, S., Altschul, C., Alvarez-Backus, E., Andrews, J., Arias, M. E., Arif, M. S., Ashkanasy, N. M., Asllani, A., … Zhou, J. (1999). Culture specific and cross-culturally generalizable implicit leadership theories. *The Leadership Quarterly*, *10*(2), 219–256. https://doi.org/10.1016/s1048-9843(99)00018-1

Diekman, A. B., & Hirnisey, L. (2007). The effect of context on the silver ceiling: A role congruity perspective on prejudiced responses. *Personality and Social Psychology Bulletin*, *33*(10), 1353–1366. https://doi.org/10.1177/0146167207303019

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dong, M., Bonnefon, J.-F., & Rahwan, I. (2024). Toward human-centered AI management: Methodological challenges and future directions. *Technovation*, *131*, 102953. https://doi.org/10.1016/j.technovation.2024.102953

Dong, M., Conway, J. R., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2026). Fears about artificial intelligence across 20 countries and six domains of application. *American Psychologist*, *81*(1), 53–67. https://doi.org/10.1037/amp0001454

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Lawrence Erlbaum Associates, Inc.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295x.109.3.573

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol. 2, pp. 458–476). Sage Publications Ltd. https://doi.org/10.4135/9781446249222.n49

Einola, K., & Khoreva, V. (2023). Best friend or broken tool? Exploring the co-existence of humans and artificial intelligence in the workplace ecosystem. *Human Resource Management*, *62*(1), 117–135. https://doi.org/10.1002/hrm.22147

Einola, K., Khoreva, V., & Tienari, J. (2024). A colleague named Max: A critical inquiry into affects when an anthropomorphised AI (ro)bot enters the workplace. *Human Relations*, *77*(11), 1620–1649. https://doi.org/10.1177/00187267231206328

Elshan, E., Zierau, N., Engel, C., Janson, A., & Leimeister, J. M. (2022). Understanding the design elements affecting user acceptance of intelligent agents: Past, present and future.

Information Systems Frontiers, *24*(3), 699–730. https://doi.org/10.1007/s10796-021-10230-9

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689

Everett, J. A. C., Claessens, S., Knöchel, T.-D., & Reinecke, M. G. (2026). Principles for understanding trust in artificial intelligence. *Nature Reviews Psychology, 5(6), 388–401*. https://doi.org/10.1038/s44159-026-00562-1

Fraune, M. R., Oisted, B. C., Sembrowski, C. E., Gates, K. A., Krupp, M. M., & Šabanović, S. (2020). Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. *Computers in Human Behavior*, *105*, 106220. https://doi.org/10.1016/j.chb.2019.106220

Georganta, E., & Ulfert, A. S. (2024). Would you trust an AI team member? Team trust in human–AI teams. *Journal of Occupational and Organizational Psychology*, *97*(3), 1212–1241. https://doi.org/10.1111/joop.12504

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

Goštautaitė, B., Liubertė, I., Parker, S. K., & Bučiūnienė, I. (2024). Can you outsmart the robot? An unexpected path to work meaningfulness. *Academy of Management Discoveries*, *10*(3), 416–437. https://doi.org/10.5465/amd.2022.0113

Goštautaitė, B., Liubertė, I., Reay, T., Bučiūnienė, I., Blaževičius, M. D., & Moniz, A. (2026). "Matilda is lazy today": How mind perceptions trigger relational job crafting with industrial robots. *Group & Organization Management, 10596011261428932*. https://doi.org/10.1177/10596011261428932

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Gu, W., Li, M., & Zhang, S. (2025). Algorithmic supervisor and employee performance. *Production and Operations Management*, *34*(10), 3101–3118. https://doi.org/10.1177/10591478251331095

Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2021). Evolving trust in robots: Specification through sequential and comparative meta-analyses. *Human Factors*, *63*(7), 1196–1229. https://doi.org/10.1177/0018720820922080

Harris-Watson, A. M., & Larson, L. E. (2026). When tech makes the team: Mind perception as a unifying framework for human and human-agent teams. *Computers in Human Behavior*, *181*, 109000. https://doi.org/10.1016/j.chb.2026.109000

Harris-Watson, A. M., Larson, L. E., Lauharatanahirun, N., DeChurch, L. A., & Contractor, N. S. (2023). Social perception in Human-AI teams: Warmth and competence predict receptivity to AI teammates. *Computers in Human Behavior*, *145*, 107765. https://doi.org/10.1016/j.chb.2023.107765

Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 5, pp. 269–298). JAI Press.

Hein, I., Cecil, J., & Lermer, E. (2024). Acceptance and motivational effect of AI-driven feedback in the workplace: An experimental study with direct replication. *Frontiers in Organizational Psychology*, *2*, 1468907. https://doi.org/10.3389/forgp.2024.1468907

Herath, D. C., Jochum, E., & Vlachos, E. (2018). An experimental study of embodied interaction and human perception of social presence for interactive robots in public settings. *IEEE Transactions on Cognitive and Developmental Systems*, *10*(4), 1096–1105. https://doi.org/10.1109/tcds.2017.2787196

Herdel, V., Kuzminykh, A., Parmet, Y., & Cauchard, J. R. (2024). Anthropomorphism and affective perception: Dimensions, measurements, and interdependencies in aerial robotics. *IEEE Transactions on Affective Computing*, *15*(3), 1463–1474. https://doi.org/10.1109/taffc.2024.3349858

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

Jago, A. S., Raveendhran, R., Fast, N., & Gratch, J. (2024). Algorithmic management diminishes status: An unintended consequence of using machines to perform social roles. *Journal of Experimental Social Psychology*, *110*, 104553. https://doi.org/10.1016/j.jesp.2023.104553

Jeon, Y. A. (2022). Let me transfer you to our AI-based manager: Impact of manager-level job titles assigned to AI-based agents on marketing outcomes. *Journal of Business Research*, *145*, 892–904. https://doi.org/10.1016/j.jbusres.2022.03.028

Ju, H., & Aral, S. (2025). Collaborating with AI agents: Field experiments on teamwork, productivity, and performance. *arXiv preprint arXiv:2503.18238*. https://doi.org/10.48550/arXiv.2503.18238

Junker, N. M., Stegmann, S., Braun, S., & Van Dick, R. (2016). The ideal and the counter-ideal follower – Advancing implicit followership theories. The *Leadership & Organization Development Journal*, *37*(8), 1205–1222. https://doi.org/10.1108/lodj-04-2015-0085

Junker, N. M., & van Dick, R. (2014). Implicit theories in organizational settings: A systematic review and research agenda of implicit leadership and followership theories. *The Leadership Quarterly*, *25*(6), 1154–1173. https://doi.org/10.1016/j.leaqua.2014.09.002

Kambhampati, S., Saldyt, L. P., Valmeekam, K., Stechly, K., Bhambri, S., Samineni, S. R., Palod, V., Biswas, U., & Kalwar, D. (2026). Position: Stop anthropomorphizing intermediate tokens as reasoning/thinking traces. *arXiv preprint arXiv:2504.09762*. https://doi.org/10.48550/arXiv.2504.09762

Kares, F., König, C. J., Bergs, R., Protzel, C., & Langer, M. (2023). Trust in hybrid human-automated decision-support. *International Journal of Selection and Assessment*, *31*(3), 388–402. https://doi.org/10.1111/ijsa.12423

Kellogg, K. C. (2022). Local adaptation without work intensification: Experimentalist governance of digital technology for mutually beneficial role reconfiguration in organizations. *Organization Science*, *33*(2), 571–599. https://doi.org/10.1287/orsc.2021.1445

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

Koban, K., & Banks, J. (2024). It feels, therefore it is: Associations between mind perception and mind ascription for social robots. *Computers in Human Behavior*, *153*, 108098. https://doi.org/10.1016/j.chb.2023.108098

Koch, A., Smith, A., Fiske, S. T., Abele, A. E., Ellemers, N., & Yzerbyt, V. (2024). Validating a brief measure of four facets of social evaluation. *Behavior Research Methods*, *56*(8), 8521–8539. https://doi.org/10.3758/s13428-024-02489-y

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371–392. https://doi.org/10.1037/a0037215

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, *137*(4), 616–642. https://doi.org/10.1037/a0023557

Kraus, J., Miller, L., Klumpp, M., Babel, F., Scholz, D., Merger, J., & Baumann, M. (2024). On the role of beliefs and trust for the intention to use service robots: An integrated trustworthiness beliefs model for robot acceptance. *International Journal of Social Robotics*, *16*(6), 1223–1246. https://doi.org/10.1007/s12369-022-00952-4

Kühne, K., Herbold, E., Bendel, O., Zhou, Y., & Fischer, M. H. (2023). "Ick bin een Berlina": Dialect proficiency impacts a robot's trustworthiness and competence evaluation. *Frontiers in Robotics and AI*, *10*, 1241519. https://doi.org/10.3389/frobt.2023.1241519

Lanz, L., Briker, R., & Gerpott, F. H. (2024). Employees adhere more to unethical instructions from human than AI supervisors: Complementing experimental evidence with machine learning. *Journal of Business Ethics*, *189*(3), 625–646. https://doi.org/10.1007/s10551-023-05393-1

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

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

Li, M., & Bitterly, T. B. (2024). How perceived lack of benevolence harms trust of artificial intelligence management. *Journal of Applied Psychology*, *109*(11), 1794–1816. https://doi.org/10.1037/apl0001200

Liao, S., Lin, L., & Chen, Q. (2023). Research on the acceptance of collaborative robots for the industry 5.0 era – the mediating effect of perceived competence and the moderating effect of robot use self-efficacy. *International Journal of Industrial Ergonomics*, *95*, 103455. https://doi.org/10.1016/j.ergon.2023.103455

Liao, S., Lin, L., Chen, Q., & Pei, H. (2024). Why not work with anthropomorphic collaborative robots? The mediation effect of perceived intelligence and the moderation effect of self-efficacy. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *34*(3), 241–260. https://doi.org/10.1002/hfm.21024

Liberman, B. E., & Golom, F. D. (2015). Think manager, think male? Heterosexuals' stereotypes of gay and lesbian managers. *Equality, Diversity and Inclusion: An International Journal*, *34*(7), 566–578. https://doi.org/10.1108/edi-01-2015-0005

Liew, T. W., & Tan, S.-M. (2021). Social cues and implications for designing expert and competent artificial agents: A systematic review. *Telematics and Informatics*, *65*, 101721. https://doi.org/10.1016/j.tele.2021.101721

Lin, W., Zhang, M., Zhang, W., & Zhang, C. (2026). The "warm–cold" effects of algorithmic management: Newcomers organizational socialization in the context of algorithm-reshaped interpersonal interaction. *Computers in Human Behavior Reports*, *22*, 101150. https://doi.org/10.1016/j.chbr.2026.101150

Liu, X., Dong, J., & Jeon, M. (2023). Robots' "Woohoo" and "Argh" can enhance users' emotional and social perceptions: An exploratory study on non-lexical vocalizations and non-linguistic sounds. *ACM Transactions on Human-Robot Interaction*, *12*(4), 1–20. https://doi.org/10.1145/3626185

Liu, X., Yi, X., & Wan, L. C. (2022). Friendly or competent? The effects of perception of robot appearance and service context on usage intention. *Annals of Tourism Research*, *92*, 103324. https://doi.org/10.1016/j.annals.2021.103324

Lord, R. G., Epitropaki, O., Foti, R. J., & Hansbrough, T. K. (2020). Implicit leadership theories, implicit followership theories, and dynamic processing of leadership information. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*(1), 49–74. https://doi.org/10.1146/annurev-orgpsych-012119-045434

Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior & Human Performance*, *34*(3), 343–378. https://doi.org/10.1016/0030-5073(84)90043-6

Lu, L., McDonald, C., Kelleher, T., Lee, S., Chung, Y. J., Mueller, S., Vielledent, M., & Yue, C. A. (2022). Measuring consumer-perceived humanness of online organizational agents. *Computers in Human Behavior*, *128*, 107092. https://doi.org/10.1016/j.chb.2021.107092

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human-autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology*, *12*, 589585. https://doi.org/10.3389/fpsyg.2021.589585

McGuire, J., & De Cremer, D. (2023). Algorithms, leadership, and morality: Why a mere human effect drives the preference for human over algorithmic leadership. *AI and Ethics*, *3*(2), 601–618. https://doi.org/10.1007/s43681-022-00192-2

McKee, K. R., Bai, X., & Fiske, S. T. (2024). Warmth and competence in human-agent cooperation. *Autonomous Agents and Multi-Agent Systems*, *38*(1), 23. https://doi.org/10.1007/s10458-024-09649-6

Mirowska, A., & Arsenyan, J. (2025). The A(I) team: Effects of human-likeness and conformity to gender stereotypes on initial trust and willingness to work with an AI teammate. *Journal of Organizational Behavior*, 1–25. https://doi.org/10.1002/job.70009

Musick, G., O'Neill, T. A., Schelble, B. G., McNeese, N. J., & Henke, J. B. (2021). What happens when humans believe their teammate is an AI? An investigation into humans teaming with autonomy. *Computers in Human Behavior*, *122*, 106852. https://doi.org/10.1016/j.chb.2021.106852

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). *Computers are social actors.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 72–78. https://doi.org/10.1145/191666.191703

Nichols, A. L., Cook, C. L., & Cottrell, C. A. (2026). What do people desire in their leaders? An affordance management approach to trait desirability across domains. *Journal of Experimental Psychology: Applied*, *32*(1), 117–137. https://doi.org/10.1037/xap0000545

Noor, N., Rao Hill, S., & Troshani, I. (2022). Artificial intelligence service agents: Role of parasocial relationship. *Journal of Computer Information Systems*, *62*(5), 1009–1023. https://doi.org/10.1080/08874417.2021.1962213

Offermann, L. R., & Coats, M. R. (2018). Implicit theories of leadership: Stability and change over two decades. *The Leadership Quarterly*, *29*(4), 513–522. https://doi.org/10.1016/j.leaqua.2017.12.003

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Paluch, R. M., & Shum, V. (2024). The non-white standard: Racial bias in perceptions of diversity, equity, and inclusion leaders. *Journal of Applied Psychology*, *109*(7), 971–986. https://doi.org/10.1037/apl0001106

Paluch, S., Tuzovic, S., Holz, H. F., Kies, A., & Jörling, M. (2022). "My colleague is a robot" – exploring frontline employees' willingness to work with collaborative service robots. *Journal of Service Management*, *33*(2), 363–388. https://doi.org/10.1108/josm-11-2020-0406

Parker, S. K., Ballard, T., Billinghurst, M., Collins, C., Dollard, M., Griffin, M. A., Johal, W., Jorritsma, K., Kowalkiewicz, M., Kyndt, E., Lusher, D., McLennan, C. L., Miller, T., Neal, A., Paterson, J. M., Vetere, F., & Walsh, T. (2026). Quality work in the future: New directions via a co-evolving sociotechnical systems perspective. *Australian Journal of Management*, *51*(2), 534–560. https://doi.org/10.1177/03128962251331813

Ponce de Leon, R., & Bailey, E. R. (2025). The preeminence of communality in the leadership preferences of followers. *Journal of Personality and Social Psychology*, *129*(2), 226–243. https://doi.org/10.1037/pspa0000437

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., & Fomina, K. (2025). Combining human and artificial intelligence: Hybrid problem-solving in organizations. *Academy of Management Review*, *50*(2), 441–464. https://doi.org/10.5465/amr.2021.0421

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

Rezaei Khavas, Z., Kotturu, M. R., Ahmadzadeh, S. R., & Robinette, P. (2024). Do humans trust robots that violate moral trust? *ACM Transactions on Human-Robot Interaction*, *13*(2), 1–30. https://doi.org/10.1145/3651992

Riso, S., Adăscăliței, D., & Oosterwijk, G. R. (2025). Implications of algorithmic management for work, employment and social dialogue: Literature review (Working Paper No. WPEF25083). Eurofound. https://www.eurofound.europa.eu/en/publications/all/implications-of-algorithmic-management-for-work-employment-and-social-dialogue-literature-review

Rousseau, D. M. (2026). Theories and models: Acting on the difference advances organizational and management scholarship. *Organizational Psychology Review*, *16*(1), 57–72. https://doi.org/10.1177/20413866251413767

Rubin, M., Li, J. Z., Zimmerman, F., Ong, D. C., Goldenberg, A., & Perry, A. (2025). Comparing the value of perceived human versus AI-generated empathy. *Nature Human Behaviour*, *9*(11), 2345–2359. https://doi.org/10.1038/s41562-025-02247-w

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179. https://doi.org/10.1016/j.jesp.2011.10.008

Ruiz-Equihua, D., Romero, J., Loureiro, S. M. C., & Ali, M. (2023). Human–robot interactions in the restaurant setting: The role of social cognition, psychological ownership and anthropomorphism. *International Journal of Contemporary Hospitality Management*, *35*(6), 1966–1985. https://doi.org/10.1108/ijchm-05-2022-0597

Rule, N. O., Bjornsdottir, R. T., Tskhay, K. O., & Ambady, N. (2016). Subtle perceptions of male sexual orientation influence occupational opportunities. *Journal of Applied Psychology*, *101*(12), 1687–1704. https://doi.org/10.1037/apl0000148

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Savela, N., Kaakinen, M., Ellonen, N., & Oksanen, A. (2021). Sharing a work team with robots: The negative effect of robot co-workers on in-group identification with the work team. *Computers in Human Behavior*, *115*, 106585. https://doi.org/10.1016/j.chb.2020.106585

Schweitzer, S., & De Cremer, D. (2024). When being managed by technology: Does algorithmic management affect perceptions of workers' creative capacities? *Academy of Management Discoveries*, *10*(3), 375–392. https://doi.org/10.5465/amd.2022.0115

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174

Siemon, D., Elshan, E., de Vreede, T., Ebel, P., & de Vreede, G. J. (2026). Beyond anthropomorphism: Social presence in human–AI collaboration processes. *Journal of Management Studies*, *63*(2), 515–560. https://doi.org/10.1111/joms.70000

Stanton, C. J., & Stevens, C. J. (2017). Don't stare at me: The impact of a humanoid robot's gaze upon trust during a cooperative human–robot visual task. *International Journal of Social Robotics*, *9*(5), 745–753. https://doi.org/10.1007/s12369-017-0422-y

Stark, J., Tröster, C., & Van Quaquebeke, N. (2026). Controlling the control condition: A critical methodological review of control conditions in experimental management research. *Journal of Management*, *52*(6), 2133–2165. https://doi.org/10.1177/01492063261424849

Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: Implications for human resource management. *Journal of Management*, *20*(2), 503–530. https://doi.org/10.1177/014920639402000210

Sy, T. (2010). What do you think of followers? Examining the content, structure, and consequences of implicit followership theories. *Organizational Behavior and Human Decision Processes*, *113*(2), 73–84. https://doi.org/10.1016/j.obhdp.2010.06.001

Sy, T., Shore, L. M., Strauss, J., Shore, T. H., Tram, S., Whiteley, P., & Ikeda-Muromachi, K. (2010). Leadership perceptions as a function of race-occupation fit: The case of Asian Americans. *Journal of Applied Psychology*, *95*(5), 902–919. https://doi.org/10.1037/a0019501

Tarafdar, M., Page, X., & Marabelli, M. (2023). Algorithms as co-workers: Human algorithm role interactions in algorithmic work. *Information Systems Journal*, *33*(2), 232–267. https://doi.org/10.1111/isj.12389

Tsai, C.-Y., Kim, J., Jin, F., Jun, M., Cheong, M., & Yammarino, F. J. (2022). Polynomial regression analysis and response surface methodology in leadership research. *The Leadership Quarterly*, *33*(1), 101592. https://doi.org/10.1016/j.leaqua.2021.101592

Tsai, C.-Y., Marshall, J. D., Choudhury, A., Serban, A., Hou, Y. T. Y., Jung, M. F., Dionne, S. D., & Yammarino, F. J. (2022). Human-robot collaboration: A multilevel and integrated leadership framework. *The Leadership Quarterly*, *33*(1), 101594. https://doi.org/10.1016/j.leaqua.2021.101594

Van Quaquebeke, N., & Gerpott, F. H. (2023). The now, new, and next of digital leadership: How artificial intelligence (AI) will take over and change leadership as we know it. *Journal of Leadership & Organizational Studies*, *30*(3), 265–275. https://doi.org/10.1177/15480518231181731

Vanneste, B. S., & Puranam, P. (2025). Artificial intelligence, trust, and perceptions of agency. *Academy of Management Review*, *50*(4), 726–744. https://doi.org/10.5465/amr.2022.0041

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383–388. https://doi.org/10.1016/j.tics.2010.05.006

Weidemann, A., & Russwinkel, N. (2021). The role of frustration in human-robot interaction - What is needed for a successful collaboration? *Frontiers in Psychology*, *12*, 640186. https://doi.org/10.3389/fpsyg.2021.640186

Wesche, J. S., Hennig, F., Kollhed, C. S., Quade, J., Kluge, S., & Sonderegger, A. (2024). People's reactions to decisions by human vs. algorithmic decision-makers: The role of explanations and type of selection tests. *European Journal of Work and Organizational Psychology*, *33*(2), 146–157. https://doi.org/10.1080/1359432x.2022.2132940

Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior*, *101*, 197–209. https://doi.org/10.1016/j.chb.2019.07.027

Wiese, E., Weis, P. P., Bigman, Y., Kapsaskis, K., & Gray, K. (2022). It's a match: Task assignment in human–robot collaboration depends on mind perception. *International Journal of Social Robotics*, *14*(1), 141–148. https://doi.org/10.1007/s12369-021-00771-z

Wildman, J. L., Nguyen, D., Thayer, A. L., Robbins-Roth, V. T., Carroll, M., Carmody, K., Ficke, C., Akib, M., & Addis, A. (2024). Trust in human-agent teams: A multilevel perspective and future research agenda. *Organizational Psychology Review*, *14*(3), 373–402. https://doi.org/10.1177/20413866241253278

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931. https://doi.org/10.1108/josm-04-2018-0119

Wolf, F. D., & Stock-Homburg, R. M. (2023). How and when can robots be team members? Three decades of research on human–robot teams. *Group & Organization Management*, *48*(6), 1666–1744. https://doi.org/10.1177/10596011221076636

Wolf, F. D., & Stock-Homburg, R. M. (2026). One size does not fit all: Mechanisms of employees' acceptance of robotic lower-level managers. *Group & Organization Management*, *51*(3), 1300–1334. https://doi.org/10.1177/10596011251313568

Wynne, K. T., & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, *19*(3), 353–374. https://doi.org/10.1080/1463922x.2016.1260181

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer-and forgive-service robots with perceived feelings. *Journal of Applied Psychology*, *106*(10), 1557–1572. https://doi.org/10.1037/apl0000834

Yam, K. C., Eng, A., & Gray, K. (2025). Machine replacement: A mind-role fit perspective. *Annual Review of Organizational Psychology and Organizational Behavior*, *12*(1), 239–267. https://doi.org/10.1146/annurev-orgpsych-030223-044504

Yam, K. C., Goh, E. Y., Fehr, R., Lee, R., Soh, H., & Gray, K. (2022). When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human. *Journal of Experimental Social Psychology*, *102*, 104360. https://doi.org/10.1016/j.jesp.2022.104360

Yang, S., Xie, W., Chen, Y., Li, Y., Jiang, H., & Zhou, W. (2025). Warmth or competence? Understanding voice shopping intentions from Human-AI interaction perspective. *Electronic Commerce Research*, *25*(6), 4625–4654. https://doi.org/10.1007/s10660-024-09859-w

Yzerbyt, V., Koch, A., Brambilla, M., Ellemers, N., Fiske, S. T., & Nicolas, G. (2025). Dimensions of stereotypes about groups. *Current Directions in Psychological Science*, *34*(6), 365–372. https://doi.org/10.1177/09637214251345489

Acknowledgements and Declarations

Acknowledgements. We thank Robert Petrossian and Anna Wichmann for their valuable assistance with the data collection. An earlier version of this paper was presented at the 2023 Annual Meeting of the Academy of Management in Boston, MA.

Ethical considerations. Ethical approval was not required for this study because it is a review of previously published research and did not involve human participants or animal subjects.

Author contributions. Christoph Daldrop: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review & editing, Project administration. Annabelle Ludwig: Investigation, Writing – original draft. Claudia Buengeler: Conceptualization, Supervision, Writing – review & editing.

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

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

Data availability. The coding data (spreadsheet of all included studies), the codebook, and the Online Supplement are available on the Open Science Framework at https://osf.io/582jn/.

Appendix

Table A1

Summary of Key Evidence on Artificial Agents in Leader Roles

Study	R: Role D: Design	Comparison	Task (competence / assertiveness)	Social-emotional (warmth / morality)	AARC evaluation (mechanism)
Hein et al. (2024)	**R:** Feedback provider **D:** Algorithmic system	Human supervisor	No significant competence difference; AI perceived as lower in power.	Lower warmth and acceptance toward the AI feedback source.	**Incongruity:** Implied lack of social presence/experience.
Höddinghaus et al. (2021)	**R:** Allocation decision agent **D:** Algorithmic leadership system	Human leader	Lower adaptability; no significant difference in perceived data-processing capacity.	Lower benevolence but higher integrity (dual moral effect).	**Mixed:** Proposed compensatory morality through integrity.
Jago et al. (2024)	**R:** Manager **D:** Algorithmic system	Human manager	Agent competence not measured; status and job complexity concern employees.	Lower interactional justice and trust for algorithmic management (Study 5 only, exploratory).	**Incongruity:** Status deficit via lower signaled job complexity (measured mediation).
Jeon (2022)	**R:** Customer-service agent **D:** AI chatbot labeled "manager"	Human vs. AI agent (S1, S3); "Manager" vs. "Representative" vs. none	Higher knowledgeability for "manager" label.	Higher likeability and trust for "manager"; some discomfort when supervising humans.	**Mixed:** Implied framing effect (status label raises competence but adds hierarchy tension).
Li and Bitterly (2024)	**R:** AI manager **D:** Algorithmic management system with persona ("Ari")	Human manager	Lower perceived competence for AI overall, not only in high-empathy contexts.	Lower benevolence and trust for AI, concentrated in high-empathy-demand contexts (integrity differences inconsistent).	**Incongruity:** Proposed lack of emotional/experience capacity.
McGuire and De Cremer (2023)	**R:** Moral leader **D:** Algorithmic decision maker	Human and humble human leaders	Lower legitimacy and trustworthiness as moral decision makers.	Lower perceived fairness (overall justice) for AI leaders.	**Incongruity:** Proposed mind-role misfit (absent moral agency).
Wolf and Stock-Homburg (2026)	**R:** Lower-level manager **D:** Physical robots (android vs. humanoid)	Human manager and between robot types	Robots-vs-human performance comparison not reported in main text; human preferred overall.	Lower relational acceptance for robots; no significant relational difference between android and humanoid robots.	**Incongruity:** Implied relational/warmth deficit for robot managers.
Yam et al. (2022)	**R:** Supervisor **D:** Physical robot (high vs. low anthropomorphism)	— (Between robot conditions)	Higher agency attribution for high-anthropomorphic robot.	Higher perceived abusiveness of high-anthropomorphic robot (indirect via agency); warmth not measured.	**Incongruity: Increased** perceived agency without increased perceived experience.

Note. The table summarizes (a) the role and design features of the artificial agent, (b) the comparison baseline, (c) evaluative outcomes along the task dimension and social-emotional dimension, and (d) the resulting evaluation of role congruity or incongruity with the principal explanatory mechanism.

Table A2

Summary of Key Evidence on Artificial Agents in Follower Roles

Study	R: Role D: Design	Comparison	Task (competence / assertiveness)	Social-emotional (warmth / morality)	AARC evaluation (mechanism)
Yam et al. (2021)	**R:** Service robot (hotel/food service) **D:** Physical robot with anthropomorphic voice, face, and name cues	Anthropomorphic vs. non-anthropomorphic; service failure vs. no failure	Anthropomorphism increased perceived agency.	Perceived experience (emotional capacity) is associated with higher satisfaction and attenuated negative reactions to failures.	**Congruity:** Experience-like cues fit follower-role expectations for empathy and service; low perceived experience removes the buffer against service failures (lower satisfaction).
Liu et al. (2022)	**R:** Service advisor **D:** Physical robot with child-like (warm) vs. adult-like (competent) appearance	2 × 2 appearance × service context (hedonic vs. utilitarian)	Competent-appearing robots are evaluated more positively in utilitarian contexts.	Warm-appearing robots receive more favorable evaluations in hedonic contexts.	**Mixed:** Acceptance depends on fit between warmth/competence cues and task context; mismatch signals role incongruity.
Ruiz-Equihua et al. (2023)	**R:** Waiter **D:** Physical service robot	Cross-sectional survey (no comparison)	Warmth and competence are associated with customer attitudes and revisit intentions, partly through psychological ownership.	Warmth and competence perceptions are positively related to psychological ownership and customer attitudes; anthropomorphism is associated with stronger relationships.	**Congruity:** Robots perceived as both warm and competent fit follower-role expectations and are associated with psychological ownership.
Lu et al. (2022)	**R:** Online service agent **D:** Algorithmic text-based chatbot vs. human agent	Human vs. chatbot interaction in live chat support	Not assessed for the agent; competence-type trust items target the organization.	Greater perceived humanness (text-based conversational style, social presence) is linked to organizational trust via perceived relational investment.	**Congruity:** For support roles, relational humanness aligns with role expectations.
Cheng et al. (2022)	**R:** Driving assistant **D:** Physical vs. virtual agent; low vs. high anthropomorphism	2 × 2 design (high vs. low anthropomorphism × physical vs. virtual agent)	Cognitive trust lower for the physically present high-anthropomorphism agent despite no significant difference in driving performance; perceived control is a user-side mediator.	Affective trust shows an anthropomorphism × physicality interaction; lower trust for the physically present high-anthropomorphism agent only marginal.	**Incongruity:** Design may exceed follower-role expectations for controllability.

Note. The table summarizes (a) the role and design features of the artificial agent, (b) the comparison baseline, (c) evaluative outcomes along the task dimension (competence/assertiveness) and social-emotional dimension (warmth/morality), and (d) the resulting evaluation of role congruity or incongruity with the principal explanatory mechanism.

Table A3

Summary of Key Evidence on Artificial Agents in Partner and Teammate Roles

Study	R: Role D: Design	Comparison	Task (competence / assertiveness)	Social-emotional (warmth / morality)	AARC evaluation (mechanism)
Einola et al. (2024)	**R:** Colleague/team member **D:** Algorithmic system framed as "Max", your (junior) colleague	Qualitative (observer: Employee vs. manager)	Employees described "Max" with low competence (e.g., "stupid," "a beginner," "too immature"); managers emphasized Max's future potential.	Managers expressed hopeful/positive affect.	**Incongruity:** Humanlike "colleague" framing in contrast with limited capability and reliability; frustration among employees.
Georganta and Ulfert (2024)	**R:** Team member **D:** Algorithmic "AI teammate"	AI vs. human teammate, across separate two-member (S1) and three-member (S2) studies	In dyads, perceived trustworthiness (incl. its ability facet) was lower for AI; triads closer to parity.	In dyads, cognitive interpersonal trust and team identification were lower with AI (exploratory); team trust and affective trust did not differ significantly.	**Incongruity:** In dyads; attenuation in triads: Another human member dilutes the mismatch.
Goštautaitė et al. (2024)	**R:** Shop-floor partner **D:** Physical industrial robots	Qualitative (no control)	Humans reported higher *own* competence when overcoming robot flaws (robot competence seen low).	Robots were mocked as "stupid" or "childish" and treated as lower-status; warmth was not measured.	**Incongruity:** Expected reliability/efficiency of the "robot partner" role violated.
Rezaei Khavas et al. (2024)	**R:** Robotic teammate (collaborative search game) **D:** Virtual agent	Moral-violation ("mean-bot") vs. performance violation ("dud-bot")	The performance-violating robot was rated lower in competence and performance but higher in reliability and overall trust than the moral-violating robot.	Moral violations produced far stronger trust loss than performance failures (MDMT moral trust: Ethical/transparent/benevolent); clear morality penalty.	**Incongruity:** Partners are expected to be benevolent; violating morality is more incongruent than being imperfect.
Liao et al. (2024)	**R:** Coworker **D:** Physical cobot	Three anthropomorphism levels: single-arm vs. two-arm vs. humanoid cobot	Anthropomorphism raised perceived intelligence, which increased negative attitudes and reduced acceptance both directly and via heightened threat.	Not assessed: No warmth/morality measure (affective reaction captured as perceived threat).	**Incongruity:** Design overshoots the role (industrial coworker), activating threat that undermines adoption.
McKee et al. (2024)	**R:** Co-player in mixed-motive games **D:** Algorithmic agents	Prosocial vs. selfish; efficient vs. "trembling-hand" inefficient	Inefficiency signaled lower competence; steady/efficient agents seen as higher competence.	Prosocial orientation associated with higher warmth; selfish orientation read as lower warmth.	**Mixed:** Prosocial/warm agents preferred; competence did not increase (and, for stated preference, was negatively related to) acceptance.
Mirowska and Arsenyan (2025)	**R:** Prospective teammate **D:** Virtual agent with varied human likeness and gender cues ("Jamie")	Study 1: Human-like vs. almost human-like vs. cartoon-like; Study 2–3: Gender (neutral/male/female) × temperament (warm/competent)	Competence was assessed only as a manipulation check and did not differ significantly across conditions.	Humanlike representation: Lower eeriness, higher initial trust (Study 1); warm female agent elicited lowest eeriness (Study 2), reversed for warm male in Study 3.	**Mixed:** Cartoon-like design triggers eeriness, lowering initial trust; gender-stereotype (in)congruity shifts eeriness; competence perceptions unchanged**.**
Wiese et al. (2022)	**R:** Collaborative task partner/decision aid **D:** Two robots differing in *experience* (emotion capacity), both agentic	"Emotional" vs. "emotionless" robot (within-person); arithmetic vs. social task assignments	Although both robots were described as equally competent, participants still assigned arithmetic tasks more to the "emotionless" robot.	No warmth/morality rating (trust is the paper's framing, not a measured scale); the "emotional" robot was preferred (chosen) for the social task — a behavioral choice, not a warmth rating.	**Mixed:** Perceived mind (experience) guided assignment; implicit schemas can misallocate tasks despite equal competence.

Note. The table summarizes (a) the role and design features of the artificial agent, (b) the comparison baseline, (c) evaluative outcomes along the task dimension (competence/assertiveness) and social-emotional dimension (warmth/morality), and (d) the resulting evaluation of role congruity or incongruity with the principal explanatory mechanism.