The Roles of Fairness and Effectiveness in Promoting Legitimacy and Cooperation with Security Robotic Authority

Xin Ye
University of Michigan
Ann Arbor, Michigan, USA
xinye@umich.edu

Lionel P. Robert Jr.
University of Michigan
Ann Arbor, Michigan, USA
lprobert@umich.edu

Abstract

Security robots increasingly assume authoritative roles, but the underlying mechanisms for why humans cooperate with them is not well understood. This study proposed and tested a cooperation model based on legitimacy theory, focusing on how distributive fairness (outcome equity) and interactional fairness (treatment equity) influence robot legitimacy and cooperation. Using a 2 × 2 online video-based experiment with 372 U.S. participants, the authors found that both fairness types promote cooperation through value alignment, with a non-significant path through obligation to obey; meanwhile, perceived effectiveness was strongly associated with both value alignment and obligation to obey. These findings extend legitimacy theory to human–robot interaction in a U.S. context, emphasizing fairness and perceived effectiveness as key to fostering cooperation and informing ethical robot design.

CCS Concepts

• Computer systems organization → Robotics; **• Human-centered computing** → Human computer interaction (HCI).

Keywords

Human–Robot Interaction, Security Robots, Robot Authority, Fairness, Legitimacy, Value Alignment, Cooperation

ACM Reference Format:
Xin Ye and Lionel P. Robert Jr.. 2026. The Roles of Fairness and Effectiveness in Promoting Legitimacy and Cooperation with Security Robotic Authority. In *Proceedings of the 21th Annual ACM/IEEE International Conference on Human Robot Interaction (HRI '26).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The delegation of authority to robots represents a significant transformation in human–technology interactions, highlighting the evolving role of robots from assistive to authoritative positions [1, 19, 25, 44, 94]. Authority is defined as the legitimate right to command obedience or make decisions that affect others [96, 120]. Security robots, for instance, have been increasingly utilized to maintain public order and deter crime in various settings such as malls, campuses, and urban areas [29, 56, 64, 68, 102]. These robots function as institutional authorities, influencing human behavior and resource access. As society increasingly empowers robots, understanding how they gain cooperation becomes critical, as their effectiveness hinges on public support to respond to their directives.

A normative framework to understand cooperation with authorities is provided by legitimacy theory, which posits that perceived fairness is central to legitimacy [50, 51, 100, 108, 115]. Legitimacy is citizens' belief that an authority has a rightful claim to power and a corresponding moral duty to obey [50]. When authorities are seen as fair, individuals are more likely to cooperate willingly, rather than out of fear of punishment, which has traditionally been a dominant approach in social control models in the United States [77, 117]. Extending punitive authority to robots poses ethical and social challenges [128]. However, research indicates that cooperation can be fostered through perceptions of legitimacy when authorities are viewed as fair, just, and aligned with societal values [50, 51, 115]. This suggests that legitimacy theory may offer a promising avenue for understanding and enhancing human–robot cooperation.

While cooperation is essential for human–robot interaction (HRI) generally, it is particularly crucial for the effective deployment of security robots. Security robots deter unwanted activity through their presence, surveillance, and ability to alert authorities to unauthorized personnel or actions [126]. Increasingly, these robots are deployed to enforce rules, monitor spaces, and issue warnings, placing them in direct contact with the public [33, 62, 102]. However, unlike law enforcement officers, security robots lack the established customs, traditions, and social foundations that underpin legitimacy in human social structures [39, 92]. Consequently, it remains uncertain whether the fairness exhibited by security robots can cultivate legitimacy and promote cooperation [85, 123], mirroring real-world deployments such as the New York Police Department's K5 pilot that was discontinued following public concerns [103]. This raises the central question: can fairness, as exercised by security robots, generate legitimacy and foster cooperation? Addressing this question is critical for ethically integrating security robots into society.

In this paper we investigated whether security robots can build legitimacy and foster cooperation through fairness, mirroring the dynamics with human authority. We proposed and tested a legitimacy model for security robots grounded in fairness. Through a 2 (Interactional Fairness) × 2 (Distributive Fairness) experiment with 372 U.S. participants, we manipulated two dimensions of fairness—distributive and interactional—in a realistic security robot scenario. Using partial least squares structural equation modeling, we examined how fairness shapes perceptions of value alignment and obligation to obey, and how these, in turn, drive cooperation. Our findings revealed that fairness enhances perceptions of value

HRI '26, Edinburgh, Scotland, UK

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

Ye, X. and Robert, L.P. (2026). **The Role of Fairness and Value Alignment in Promoting Cooperation with Security Robotic Authority**, *Proceedings of the 20th ACM/IEEE International Conference on Human-Robot Interaction* (**HRI 2026**), (Acceptance rate: 23%), March 16-19, 2025 - Edinburgh, Scotland, UK.

alignment but does not significantly increase obligation to obey. Interestingly, perceived effectiveness shows the strongest associations with both value alignment and obligation to obey. This suggests that both fairness and perceived effectiveness are key pathways to cooperation. Fairness promotes cooperation through value alignment, with a non-significant path through a sense of duty, highlighting a key difference between human and robotic authority.

Our findings contribute to the literature in multiple ways. Theoretically, we extend legitimacy theory to HRI, highlighting both similarities and differences. In doing so, we establish boundaries to legitimacy theory's applicability in robotic contexts. First, they reveal that fairness perceptions of robots mainly promote value alignment rather than create a general sense of duty to obey. Second, they show that the obligation to obey remains an important predictor of cooperative behavior even when it is not directly caused by the robot's fairness signals. Third, they emphasize that legitimacy judgments are influenced by broader institutional views (such as perceptions of law enforcement) and by practical evaluations of robot effectiveness. Overall, these findings suggest design and policy considerations for deploying security robots: to encourage cooperation, robots should be designed and operated to communicate fair treatment and outcomes that foster value alignment, and institutional practices—including human policing—and proven effectiveness should be taken into account to build public trust.

2 Background

2.1 What is Authority

2.1.1 Authority in Human–Robot Interaction. Authority is commonly understood as the socially recognized right to command and expect obedience [26, 120]. Authority spans multiple levels in human societies. At the macro level, political authority is exercised through governments, legal systems, and state institutions that regulate collective life [113, 120]. In Max Weber's classic formulation, power is the ability to impose one's will on others, whereas authority is power regarded as just and appropriate: “the probability that a command with a given specific content will be obeyed by a given group of persons” [120, p. 152]. At the meso level, authority operates within organizations, communities, and social groups [7, 96], where it is often defined as the power to make and influence decisions that guide the actions of others [96]. At the micro level, authority is enacted in everyday life, from family dynamics to professional hierarchies, rooted in interpersonal power relations [35].

As robots become increasingly integrated into public and organizational life, questions have emerged about whether and how they can exercise authority [1, 25, 94]. In HRI research, robot authority is understood as the capacity of robots to issue directives and influence human behavior [25]. Saunderson and Nejat [94] highlighted two main forms that scholars have examined: formal authority, when robots are assigned a specific identity or role associated with authority within a given context [1, 2, 5]; and real authority, when robots exercise direct control over human actions [25, 75].

Research on authority in HRI has developed along two primary directions. The first concerns whether robots can actually embody authority and elicit obedience. Empirical studies demonstrate that people sometimes obey robots even in tedious or uncomfortable tasks, suggesting that robots can exert compliance pressures similar to human authorities [25, 37, 38, 118]. Drawing on Milgram's paradigm, Geiskkovitch et al. [38] instructed participants to persist in monotonous tasks at the request of a robot. Although obedience rates were lower than with human experimenters, nearly half of participants still complied until the end. Similarly, Haring et al. [45] showed that participants often practiced longer than they desired when prompted by robotic coaches, indicating that robots can exert meaningful social pressure. These findings suggest that, under certain conditions, robots can function as authority figures capable of shaping human behavior.

The second line of research examined the design features and contextual cues that shape compliance with robotic authority. Factors such as embodiment [6], group membership [95], human-likeness [45], and strategy (reward versus punishment) have been examined to see how they influence compliance. For example, Sembroski et al. [95] found that participants followed a robot more when it was seen as in-group. Saunderson and Nejat [94] found the robot was more persuasive when giving rewards compared to punishments. Bainbridge et al. [6] found human compliance between the physical and video robots was similar for simple tasks.

Taken together, robotic authority has emerged as an important research focus. Findings suggest that robots can indeed function as authority figures and elicit obedience, while design features play a role in shaping how robotic authority influences compliance.

2.1.2 Authority in Security Robots. Security robots represent one of the most visible and socially consequential forms of robotic authority [29, 56, 57, 70, 86]. They are commonly defined as robots "deployed to prevent unwanted activities through their presence, surveillance, and ability to notify authorities of unauthorized personnel or actions" [126, p. 1]. Unlike service robots that primarily assist with tasks, security robots are explicitly designed to monitor, regulate, and enforce social order in public and organizational spaces. High-profile examples include Knightscope K5 robots deployed by the New York Police Department in Times Square and railway stations, where they are tasked with deterring misconduct, issuing warnings, and in some cases reporting or escalating incidents to human officers [102]. Their presence in highly visible and sensitive environments positions them as robots that visibly enact legal authority in the public sphere.

What makes security robots distinctive among robotic authorities is the nature of the authority they exercise. Whereas most discussions of robotic authority focus on the interpersonal level—such as robots directing individuals in tasks or experiments [25, 37, 75, 94]—the authority of security robots is closely tied to law enforcement and public safety [68], representing legal authority at the societal level delegated from human officers to robots. However, compared to human officers, whose authority is grounded in longstanding institutional histories, legal mandates, and established professional norms [50, 71], the authority of security robots can be fragile and contested. Lacking institutional histories and legal status, security robots introduce uncertainty about whether people will cooperate with their directives. Unsurprisingly, their deployment has provoked public controversy [68, 103]. Ethical and societal concerns have been raised about their use, ranging from surveillance in enforcement to the potential reinforcement of inequities in public safety [85, 123]. These concerns make it imperative to understand how security robots should be designed to ensure their exercise of authority is both legitimate and socially accepted.

At the same time, recent HRI research has begun to investigate security robots directly, examining whether and how they can elicit behavioral compliance [1, 2, 10, 59, 73] and foster perceptual acceptance [36, 48, 66, 67, 124–126, 128, 129]. Compliance can be defined as people's behavioral adherence to issued directives [108], whereas acceptance refers to people's intention or willingness to use and rely on security robots [28]. Studies on compliance indicate that the mere presence of a security robot can increase adherence to rules [2, 59, 73]—for example, by making people more likely to adhere to mask-wearing policies during COVID-19 [59]. Cross-cultural work further shows that compliance varies by national context: Chinese participants residing in the United States exhibited the highest compliance with peacekeeping robots, whereas Americans living in China showed the lowest [10]. Beyond behavioral outcomes, a larger body of research has explored perceptual acceptance, focusing on the influence of human attributes [31, 36, 48, 91], robot design features [65, 67, 124, 125], and contextual factors [60, 68, 125]. A recent paper reviewed human–security robot interaction literature in the field and highlighted the diverse factors that have been explored to shape human performance, perceptions and acceptance of security robots [126]. However, to be notice, cooperation—defined as people's willingness to "assist the agent with crime prevention, such as by reporting crime or calling for help" [100, p. 466]—has not yet been systematically examined in this context.

Taken together, this growing body of work demonstrates that security robots have become an increasingly important focus of HRI research. Yet, to fully understand how people respond to security robotic authority, it is essential to move beyond compliance and acceptance and systematically investigate cooperation.

2.2 Tyler's Legitimacy theory

A central question in the study of authority is not only what authority is, but why people choose to follow it. Classic psychological research has demonstrated that individuals often comply with authority figures, sometimes even against their own preferences or values [9]. Stanley Milgram's [72] obedience experiments famously showed that ordinary people could be induced to administer what they believed were harmful electric shocks when instructed by an authority figure. These findings underscore that authority is not merely a structural position but also a psychological force shaping compliance and cooperation [55].

Building on this, Tom Tyler [108] proposed the legitimacy model to explain why people cooperate with the police and obey the law. Since then, it has become one of the most influential frameworks for understanding why people follow authority [50]. His model emphasizes that legitimacy serves as the key predictor of people's cooperation: individuals comply and cooperate with legal authorities when they ascribe legitimacy to them [100, 108, 110, 115]. At the same time, legitimacy is rooted in fairness—people are more likely to comply when they perceive authority as fair, just, and aligned with shared values [61, 106, 113, 115, 117]. This approach stands in contrast to traditional deterrence models, which posit that compliance arises from fear of punishment or negative consequences [27, 83]. Instead, the legitimacy model highlights perceptions of fairness and legitimacy as the foundation of enduring authority.

Tyler's legitimacy model emphasizes process-based fairness when citizens encounter and interact with authorities (often referred to as procedural justice) [61] , and identifies two central criteria for evaluating it: (1) the quality of decision-making, which concerns whether authorities explain their reasoning during the interaction and allow individuals the opportunity to express their views; and (2) the quality of interpersonal treatment, which concerns whether authorities show respect, courtesy, and helpfulness [97, 100, 106, 116]. Therefore, to evaluate process-based fairness, people judge whether they were treated with dignity, approached in a friendly manner, and given clear explanations for decisions [113]. Notably, what Tyler termed the two-dimensional "process-based fairness" parallels what organizational scholars have classified as "interactional fairness" [40]. Interactional fairness is defined as encompassing both interpersonal fairness—"the degree to which employees are treated with respect and dignity"—and informational fairness—"the degree to which information is provided to help employees understand the processes that determine outcomes" [90, p. 547]. In this paper, we adopt this terminology and refer to process-based fairness as interactional fairness, encompassing its two dimensions of interpersonal and informational fairness.

Subsequent legitimacy research extended Tyler's model by incorporating distributive justice—whether authorities allocate outcomes such as arrests, citations, protection, and services fairly across social groups—into the framework [47, 71, 121]. In this view, people not only assess the fairness of treatment but also evaluate the fairness of outcomes when forming judgments about the legitimacy of authority. Although distributive justice contributes to legitimacy and cooperation, it has generally received less attention, and its effects are often found to be weaker than those of procedural justice [71, 88, 101, 121]. Interestingly, this concept closely parallels the notion of distributive fairness in organizational research [40], which refers to "fairness with respect to the allocation of outcomes such as pay and other resources" [90, p. 547].

In practice, Tyler's legitimacy model has profoundly shaped institutional practices within the U.S. legal system [114]. Traditionally, U.S. policing was guided by deterrence-based approaches, relying on strategies such as expanding police presence to discourage crime and imposing harsher punishments on convicted offenders to deter future offenses [54, 133]. Over the last two decades, however, the legitimacy model has gained increasing influence, shifting reform efforts toward emphasizing the fairness of authority [50, 78, 114]. The President's Task Force on 21st Century Policing explicitly identified legitimacy as the foundational pillar of policing and promoted process-based justice as the key pathway to achieving it [79].

Yet, our understanding of whether legitimacy theory can be applied to enhance cooperation with robots remains unexplored, yet the topic seems of vital importance to security robots. Prior work has only investigated isolated design features, such as politeness or communication style, and explored their influence on perceptions and acceptance of security robots [48, 49, 63]. While these studies can be interpreted as engaging with subdimensions of interactional fairness, especially interpersonal fairness, they remain fragmented and lack a unifying theoretical framework that explains why these features matter. More importantly, they have not been linked to

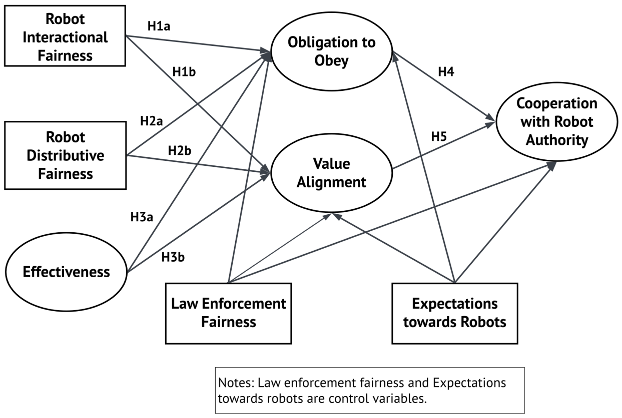

Figure 1: Tyler's legitimacy model

the broader questions of legitimacy and cooperation that are central to authority research. At the same time, existing HRI fairness research—such as work on teammate fairness [17, 18], group dynamics [21, 22], and fairness-aware behaviors [16]—has not examined diverse fairness dimensions or linked them to legitimacy or cooperation with robotic authority. To address these gaps, the present study drew on fairness and legitimacy theory to experimentally manipulate distributive and interactional fairness in a security robot scenario to investigate how different forms of fairness influence legitimacy and cooperation with security robotic authority.

3 Research Model and Hypotheses

We propose a legitimacy model for security robots in which two forms of fairness—interactional and distributive fairness—together with perceived effectiveness, shape two legitimacy-relevant constructs—value alignment and obligation to obey—that, in turn, drive cooperation with the robot authority (Figure 1).

Interactional fairness concerns the fairness of how individuals are treated in encounters with authority. Extensive legitimacy research, conducted across the United States [100, 107–109, 116] and more than seventy other countries [11, 12, 47, 80, 81], consistently shows that interactional fairness is the strongest and most reliable predictor of perceived legitimacy. People evaluate their experiences with authorities based on whether they were treated fairly and form judgments about the legitimacy of those authorities.

But what is legitimacy? Scholars distinguish it into two primary dimensions: obligation to obey and normative alignment [14, 43, 47, 51, 52, 105]. Tyler [110] argued that legitimacy arises from a sense of obligation that transcends material incentives or self-interest. Rather than stemming from an authority's capacity to reward or punish, legitimacy is grounded in qualities that lead people to view the authority as entitled to make decisions and command obedience. Tyler later expanded this conception to include moral alignment—the perception that the authority shares core values with the public [111, 112]. Such alignment provides a normative foundation for institutional authority [51]. In this way, legitimacy represents a normative judgment about the rightfulness of authority, encompassing not only a felt obligation to obey but also perceptions of value congruence with that authority [50, 112].

Building on this framework, we extend the legitimacy model to robotic authority. When security robots operate as institutional authorities in public spaces, interactional fairness is expected to play a pivotal role in shaping perceptions of legitimacy. People are likely to scrutinize whether these robots communicate clearly, deliver instructions respectfully, and treat individuals with dignity during encounters such as monitoring, issuing warnings, or enforcing rules. These evaluations, in turn, shape whether security robots—like human officers—are regarded as legitimate authorities:

> ***H1:*** *Greater* ***interactional fairness*** *increases (a)* ***obligation to obey*** *and (b)* ***value alignment*** *with security robots.*

Distributive fairness focuses on the equity of results and the consistency with which authorities apply rules and sanctions. In human authority contexts,it signals impartiality and adherence to shared standards of justice, thereby reinforcing legitimacy [30, 71]. While interactional fairness has been consistently identified as the strongest predictor of legitimacy, distributive justice remains an important factor shaping whether people regard authorities as legitimate [46, 88, 97]. Extending this logic to robotic authority, we expect distributive fairness to similarly influence legitimacy perceptions. When security robots enforce rules or make decisions—such as issuing warnings, reporting incidents, or restricting access—people are likely to evaluate not only how they are treated but also whether the outcomes are fair and consistent. If security robots distribute sanctions or benefits equitably, individuals may infer that the robots embody shared principles of justice, thereby strengthening both perceived value alignment and the felt obligation to obey:

> ***H2:*** *Greater* ***distributive fairness*** *increases (a)* ***obligation to obey*** *and (b)* ***value alignment*** *with security robots.*

Perceived effectiveness represents an instrumental pathway to legitimacy, focusing on whether authorities are seen as capable of achieving desired outcomes [8, 47, 52, 100]. Unlike fairness-based judgments, which are normative in nature, effectiveness reflects a utilitarian calculation [47, 58, 101]: people are more likely to regard authorities as legitimate when they believe those authorities can "get the job done." Research on policing consistently finds that perceptions of effectiveness—such as the ability to deter crime or respond promptly to incidents—contribute significantly to legitimacy, sometimes even outweighing procedural fairness in particular socio-political contexts [58, 98, 101]. In China and South Korea, for example, police effectiveness emerged as the strongest predictor of legitimacy, suggesting that citizens evaluate legitimacy not only through fair processes but also through whether authorities demonstrate competence in fulfilling their responsibilities [58, 98, 99]. When it comes to security robots, we expect perceived effectiveness to play a similarly important role, because it signals that the robot is fulfilling its rightful role in providing safety, thereby shaping people's perceptions of legitimacy.

> ***H3: Perceived effectiveness*** *of the security robot is positively related to (a)* ***obligation to obey*** *and (b)* ***value alignment****.*

In Tyler's model, cooperation represents a broader form of support that goes beyond obedience [100, 112]. Legitimacy theory posits that when people view an authority as legitimate, they feel an intrinsic obligation to obey and regard the authority's goals as worthy of support [50, 108, 113]. This sense of legitimacy motivates not only passive compliance but also active cooperation, such as reporting crimes, providing information, or testifying in court [106, 107, 109]. We therefore hypothesize that if individuals perceive robot authority as legitimate, they will be more willing to voluntarily cooperate in achieving institutional objectives.

> ***H4: Obligation to obey*** *is positively related to* ***cooperation*** *with security robots.*

At the same time, legitimacy also rests on perceptions of value alignment—the belief that an authority shares and upholds people's core values and principles [51, 112]. When individuals perceive such normative alignment, they view the authority's goals as morally justified and socially beneficial, and are therefore more willing to support that authority [51, 74, 76, 87, 111, 112]. Applied to robotic authority, if people see security robots as value-aligned, they should go beyond compliance and actively cooperate to promote collective safety and order. Thus, we hypothesize:

> ***H5: Value alignment*** *is positively related to* ***cooperation*** *with security robots.*

4 Methods

To test our hypotheses, we conducted a 2 (interactional fairness: low vs. high) × 2 (distributive fairness: low vs. high) between-subjects experiment. Participants were randomly assigned to one of four conditions, where they viewed a real-world video clip of a human–security robot interaction and then responded to a survey about the interaction. The study protocol was reviewed and approved by the university's institutional review board. The study was not pre-registered. All study materials (video clips, vignettes, and measures) are available in the supplementary material.

4.1 Participants

A total of 399 participants were recruited through Prolific. 27 cases were excluded prior to final analysis: seven failed an instructed-response attention-check question, eight failed a video memory-check designed to ensure careful viewing of the stimulus, and twelve exhibited excessively long completion times (over one hour). The final valid sample consisted of 372 participants (180 women, 182 men, 10 identifying as other), ranging in age from 19 to 83 years (M = 47, SD = 16). Participants were randomly assigned to one of four experimental conditions, with 93 participants per condition. All participants completed the study and received $8 USD in compensation. Mean completion time was 16 minutes. Eligibility criteria required participants to be at least 18 years old, fluent in English, and residing in the United States. Demographic characteristics were as follows: 20% resided in the Midwest, 16% in the Northeast, 43% in the South, and 21% in the West; 60% identified as White, 15% as Hispanic or Latino American, 14% as Black or African American, 7% as Asian, and 4% as other or multiracial. A post-hoc power analysis (G*Power 3.1; $f^2 = 0.05$, N = 372, predictors = 5) confirmed adequate power (0.93) to detect small-to-medium effects.

Figure 2: Screenshots of videos

4.2 Scenario, Task, and Procedure

Participants first completed a pre-questionnaire to provide demographic information. They were then randomly assigned to one of the experimental conditions and watched a real-world video (see Fig. 2 for example screenshots) depicting an interaction between a security robot (Knightscope K5) and a citizen, Tom, who was attempting to park his car for a scheduled appointment at a company. In the video, the robot approaches Tom's vehicle while patrolling the parking lot and denies entry, followed by a brief exchange of dialogue. After the video, participants were presented with a short vignette describing subsequent events at the gate, in which Tom remained in his car as two other vehicles attempted to enter. Finally, participants completed a post-questionnaire about the interaction. Participants were free to withdraw from the study at any time.

4.3 Experimental Design

We employed a 2 (interactional fairness: low vs. high) × 2 (distributive fairness: low vs. high) between-subjects design. Participants were randomly assigned to one of four conditions and took part in the study as third-party observers. Each participant first viewed a real-world interaction video manipulating interactional fairness and then read a post-video vignette manipulating distributive fairness. This procedure yielded four distinct experimental conditions. Such mixed-modality manipulation has been widely used [4, 15, 41, 130].

4.3.1 Interactional Fairness. Interactional fairness was manipulated through its sub-dimension informational fairness, operationalized as the presence or absence of an explanation. Participants watched one of two videos depicting a security robot interacting with a citizen. The videos were identical in setting, dialogue, and outcome, except that in the high-interactional fairness condition the robot explained why access was denied, whereas in the low-interactional fairness condition the robot did not explain.

4.3.2 Distributive Fairness. Distributive fairness was manipulated based on whether the security robot delivered equal or unequal outcomes. In the high distributive fairness condition, the robot denied access not only to the citizen in the video but also to two additional unregistered vehicles, ensuring equal outcomes. In the low distributive fairness condition, the robot denied access to the citizen but granted temporary entry to two unregistered high-end vehicles, creating unequal outcomes.

4.4 Measures

4.4.1 Manipulation Check Measures. Perceived informational fairness was measured using a 5-item scale adapted from [23]. The scale was reliable (Cronbach's α = .90). Perceived distributive fairness was assessed using a 4-item scale adapted from [71] (α = .95).

4.4.2 Control Variables. We measured participants' age, gender, ethnicity, and region. Expectations toward security robots used a 4-item scale from [132] (α = .97), and perceptions of law enforcement officers' fairness used a 4-item scale from [71] (α = .95).

4.4.3 Dependent Variables. Perceived effectiveness (4 items; [98]; α = .87), obligation to obey (3 items; [71]; α = .93), value alignment (3 items; [98]; α = .94), and willingness to cooperate (3 items; [98]; α = .93) were assessed in the post-questionnaire.

5 Results

In this section, we report the study results. Main analyses were conducted with PLS-SEM in SmartPLS 4.1 [89], and manipulation checks with R. Figure 3 displays the detailed results of the final model, including standardized path coefficients (β), their corresponding standard errors (SE), and the explained variance of the endogenous constructs (R^2). Demographic factors (age, gender, region, ethnicity) were tested as controls but were non-significant and thus excluded from the final model.

5.1 Manipulation Check

To ensure the effectiveness of the experimental manipulations, independent-samples t-tests were conducted. Assumptions including independent samples, equality of population variances, and normality were checked and satisfied. For informational fairness, participants in the high condition (M = 5.47, SD = 0.87) reported significantly higher perceived informational fairness than those in the low condition (M = 4.54, SD = 1.06), $t(370)$ = -6.76, $p < .001$, Cohen's d = 0.70, 95% CI [0.49, 0.91].

Similarly, for distributive fairness, participants in the high condition (M = 5.66, SD = 0.93) reported significantly higher perceived distributive fairness than those in the low condition (M = 3.63, SD = 1.13), $t(370)$ = -13.43, $p < .001$, Cohen's d = 1.39, 95% CI [1.16, 1.62]. These results confirm that the manipulations were successful.

5.2 Measurement Validity

Factor analysis was conducted to evaluate the measurement model. All items loaded strongly on their intended constructs ($\geq$ 0.70), with no cross-loadings above 0.40, except for the first item of the perceived effectiveness scale ("The robot responds swiftly to calls for need"), which was removed. Convergent validity was confirmed using the Fornell–Larcker criterion [34], with all average variance extracted (AVE) values—willingness to cooperate (0.88), perceived effectiveness (0.80), value alignment (0.90), and obligation to obey (0.87)—exceeding the 0.50 threshold. Discriminant validity was further supported, as the square roots of the AVEs were consistently greater than the inter-construct correlations (Table 1). Finally, internal consistency was established through internal composite reliability (ICR), with values of 0.96 (willingness to cooperate), 0.92 (perceived effectiveness), 0.96 (value alignment), and 0.95 (obligation to obey) all well above the 0.70 threshold.

5.3 Hypothesis Testing

Hypothesis 1 predicted that distributive fairness would positively influence perceptions of value alignment and obligation to obey. The results partially supported this hypothesis. Distributive fairness condition had a significant positive effect on value alignment ($\beta = 0.25, SE = 0.04, p < .001$), but its effect on obligation to obey was not significant ($\beta = 0.07, SE = 0.04, p = .08$). Thus, H1a was supported, whereas H1b was not supported.

Hypothesis 2 proposed that informational fairness would positively influence value alignment and obligation to obey. Consistent with expectations, informational fairness significantly increased value alignment ($\beta = 0.11, SE = 0.03, p = .001$), but its effect on obligation to obey was not significant ($\beta = -0.07, SE = 0.04, p = .08$). Therefore, H2a was supported, but H2b was not supported.

Hypothesis 3 stated that perceived effectiveness of security robots would positively relate to value alignment and obligation to obey. The results supported this hypothesis: effectiveness significantly predicted both value alignment ($\beta = 0.57, SE = 0.04, p < .001$) and obligation to obey ($\beta = 0.58, SE = 0.06, p < .001$).

Hypothesis 4 posited that obligation to obey would increase cooperation, whereas Hypothesis 5 predicted the same effect for value alignment. The results showed that both value alignment ($\beta = 0.16, SE = 0.06, p = .004$) and obligation to obey ($\beta = 0.24, SE = 0.06, p < .001$) significantly predicted cooperation, fully supporting H4 and H5. Taken together, these findings provide partial support for our proposed theoretical model. Table 2 summarizes the results.

6 Discussion

In this study, we aimed to understand why people cooperate with security robotic authority by applying the lens of legitimacy theory. We investigated whether robot fairness can generate legitimacy and, in turn, foster cooperation, based on a U.S. sample and online video interactions. Our results found that interactional and distributive fairness improved perceptions of value alignment and promote cooperation but did not significantly strengthen the obligation to obey. In contrast, perceived effectiveness robustly predicted both value alignment and obligation, and promoted cooperation.

6.1 Contributions

First, this paper extends legitimacy theory, traditionally applied to human authority [26, 113, 120], into the realm of HRI, demonstrating its relevance in explaining how individuals perceive and respond to robotic authority. Specifically, we demonstrate that the perceived interactional and distributive fairness of a security robot promotes cooperation primarily through value alignment; while perceived effectiveness through both aligned values and a sense of duty. This is significant because while prior work has shown that robots can elicit obedience [25, 37, 38, 131], our work helps explain why humans choose to cooperate.

Second, our research reveals a critical divergence from traditional legitimacy theory [106, 107, 109]. While fairness promotes value alignment, it does not significantly lead to a sense of obligation to obey the robot. Instead, the obligation to obey appears to operate as an independent pathway to cooperation. This contrasts with human authority, where fairness often fosters both value alignment and a sense of duty or obligation [61, 100]. This pattern aligns with

Table 1: Descriptive statistics and correlation matrix

Variable	M	SD	1	2	3	4	5	6	7	8
1. Cooperation	4.54	1.69	(0.94)							
2. Distributive Fairness	0.50	0.50	0.08	(N/A)						
3. Law Enforcement Distributive Fairness	4.60	1.47	0.28**	-0.03	(N/A)					
4. Effectiveness	4.77	1.30	0.56**	0.33**	0.19**	(0.89)				
5. Informational Fairness	0.50	0.50	-0.03	0.00	-0.01	0.09	(N/A)			
6. Value Alignment	4.27	1.61	0.50**	0.43**	0.24**	0.74**	0.16**	(0.95)		
7. Obligation to Obey	4.87	1.50	0.52**	0.26**	0.29**	0.65**	-0.02	0.63**	(0.94)	
8. Expectations	4.20	1.48	0.59**	-0.04	0.27**	0.50**	-0.00	0.41**	0.38**	(N/A)

Notes: 1. N = 372. Significance of correlations: *p < .05; **p < .01.
2. Values on the diagonals (in parentheses) represent the square root of the average variance extracted (AVE).
3. Experimental conditions "Distributive Fairness" and "Informational Fairness" were coded as 0 = low and 1 = high.
4. Control variables "Law Enforcement Distributive Fairness" and "Expectations toward Robots" were included as observed composite scores (mean values of the respective items).
5. Values are rounded to two decimals; '-0.00' indicates a small negative value (>-0.005).

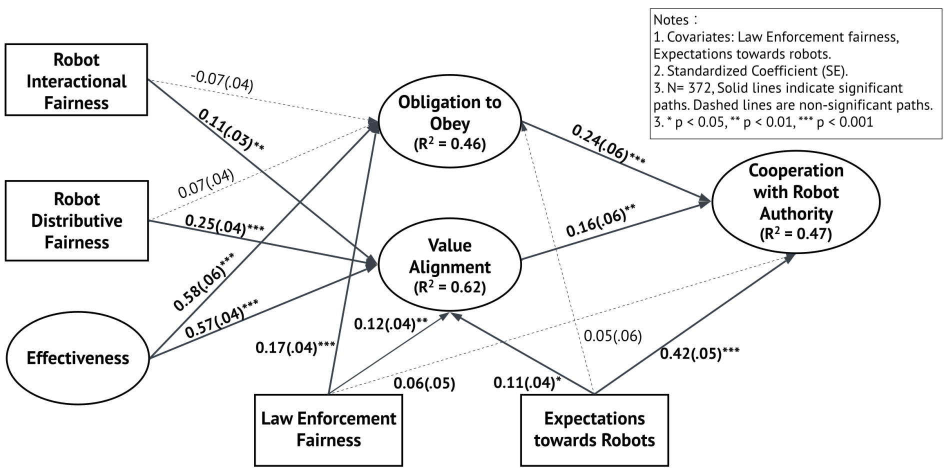

Figure 3: Results of PLS analysis

Table 2: Results of Hypothesis Testing

Hypothesis	**Results**
H1) Greater interactional fairness increases (a) obligation to obey and (b) value alignment with security robots.	Partially Supported
H2) Greater distributive fairness increases (a) obligation to obey and (b) value alignment with security robots.	Partially Supported
H3) Perceived effectiveness of the security robot is positively related to (a) obligation to obey and (b) value alignment.	Supported
H4) Obligation to obey is positively related to cooperation with security robots.	Supported
H5) Value alignment is positively related to cooperation with security robots.	Supported

[38]'s findings that robots elicit weaker obedience than humans, suggesting potential inherent limits on robots' ability to generate a sense of duty. The findings strengthen the importance of refining and applying legitimacy theory to specify HRI context. At the same time, the finding that fairness promotes cooperation through shared values helps explain why prior studies identified design features—such as politeness [48, 49]—as influential in HRI: such features may have signaled fairness and thereby fostered value alignment. Our results also highlight and expand the importance of fairness in HRI beyond team dynamics [17, 18, 21] to the domain of robotic authority, deepening fairness's multifaceted treatment and linking it to robot legitimacy. Future research could investigate the independent obligation-based pathway and the conditions under which instilling a sense of obligation to obey in robots is effective.

Third, the findings highlight the significance of external perceptions in shaping attitudes toward security robots. Our findings suggest that perceptions of fairness in law enforcement have a positive influence on value alignment and the obligation to obey the robot. This demonstrates that legitimacy judgments about robots are shaped by broader institutional contexts and instrumental assessments, extending prior HRI work on contextual and design cues that affect compliance (e.g., embodiment [6], group membership

[95]). This suggests that public trust in broader institutional authorities can spill over and influence perceptions of robotic authority. This underscores the need for a holistic approach to deploying security robots, considering not only the robot's design but also the broader social and institutional context in which it operates.

Furthermore, the study showed that perceived effectiveness of the security robot more strongly promotes value alignment and obligation to obey the robot, highlighting the importance of robots that are both fair and efficacious. This pattern differs from findings in human law-enforcement contexts in the U.S., where fairness has consistently emerged as the stronger predictor of legitimacy than effectiveness [100, 116]. This divergence highlights the unique nature of robotic authority and points to the need for the field to prioritize competence in the development of security robots. It may also help explain prior findings that design features such as reliability [67] increase security robot acceptance, likely because they serve as cues of effectiveness.

6.2 Implications for Future Research

First, researchers should investigate design features that foster value alignment. Given that fairness perceptions primarily drive cooperation through value alignment, HRI design work should prioritize interactional and distributive cues that communicate respect, neutrality, and equitable outcomes. Experimental work should further identify which specific communicative behaviors [48, 49], decision-logic explanations, or outcome distributions most effectively produce value alignment across diverse populations.

Second, researchers should test boundary conditions and generalizability by assessing whether the primacy of value alignment holds across different robot roles (e.g., care robots [84], education robots [3]), task stakes (low vs. high risk) [130], scenarios (e.g., security-initiated vs. citizen-initiated interactions) [127], and interaction modalities (autonomous vs. teleoperated) [104]. Mixed-methods and field designs will be valuable for external validity.

Third, researchers could explore pathways to a sense of obligation to obey. Because robot fairness did not significantly generate an obligation to obey, researchers should investigate alternative sources of felt obligation, such as legal mandates [53], anthropomorphic cues [82], or explicit authority labeling [59], and how these sources interact with fairness to influence behavior.

Fourth, researchers should examine cross-institutional legitimacy dynamics. Our findings suggest that legitimacy perceptions transfer across related human and robotic institutions. Future work should investigate when and why trust or distrust in human institutions (e.g., police) amplifies or attenuates the effects of robot legitimacy, including the moderating roles of prior experience, cultural context [24], and threat salience [69].

6.3 Implications for Design

The findings also offer key design implications for security robots, emphasizing fairness and perceived effectiveness.

1. Prioritize fairness signaling. This includes designing robot behaviors that convey interactional fairness, such as polite, respectful language; clear, nonjudgmental explanations for requests; and attentive listening cues (e.g., prompts that acknowledge user perspectives), as well as communicating distributive fairness through transparent, consistent, and justifiable allocation of benefits, burdens, or sanctions (e.g., explicit rules, visible decision criteria, consistent application across people).

2. Emphasize robot competence. Incorporate cues that signal reliable, accurate, and timely task performance (e.g., clear status displays, confirmations of successful actions) to strengthen perceived effectiveness and support cooperation.

3. Align robot goals with societal values. Security robots should be designed to reflect the values of the communities they serve. This could involve engaging with community stakeholders during the design process to ensure that the robot's objectives and behaviors align with local norms [58, 59]. Values can also be made salient and relatable during interaction by providing short, user-facing rationales that connect robot actions to collective goods.

4. Leverage institutional context thoughtfully. This could include coordinating robot behaviors and messaging with organizational partners (e.g., security staff) [93] so robot actions align with broader institutional norms and public perceptions of human authorities. When operating in environments with low trust in local institutions, emphasize transparency, neutrality, and community engagement features to mitigate negative transfer.

5. Design for fairness across diverse populations. This could involve testing interactions with diverse user groups to ensure fairness cues are perceived consistently and incorporating culturally sensitive language, multilingual options, and adaptable interaction styles. Use participatory design [119] and community consultation [13] to surface local norms about fairness and authority.

6.4 Limitations

This study has several limitations. First, our sample consisted of U.S. participants. Prior research on legitimacy suggests that the relative importance of antecedents varies across sociocultural contexts [58, 98, 101], so the cooperation pathways we identified may be context-specific. Future research should employ cross-national designs to test how the legitimacy model generalizes and shifts across cultures. Second, we focused on security robots as one form of robotic authority, yet authority can manifest in diverse domains (e.g., healthcare triage [42], workplace supervision [122], education [20]) with distinct stakes, beneficiaries, and accountability structures. Subsequent work should test the boundary conditions of the proposed model across these contexts. Third, our experiment relied on a video-based, third-party observer paradigm. While this method has been validated as producing results comparable to in-person approaches [32], it still poses ecological validity limitations that may not fully capture behavioral dynamics or emotional responses in in-situ encounters. Future in-person and longitudinal studies are encouraged to further explore these aspects.

7 Conclusion

This study advances HRI scholarship by integrating legitimacy theory into the study of robotic authority, disentangling internalized and deference-based pathways to cooperation, and highlighting the role of broader institutional and instrumental cues. These contributions open specific directions for experimental, design-oriented, and cross-disciplinary research aimed at understanding and shaping cooperative relations between people and security robots.

References

[1] Siddharth Agrawal and Mary-Anne Williams. 2017. Robot authority and human obedience: A study of human behaviour using a robot security guard. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction*. 57–58.

[2] Siddharth Agrawal and Mary-Anne Williams. 2018. Would you obey an aggressive robot: A human-robot interaction field study. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 240–246.

[3] Verónica Ahumada-Newhart and Jacquelynne S Eccles. 2025. Robots for learning: An exploration of teacher roles, perceptions, and challenges in robot-mediated learning. *Frontiers in Robotics and AI* 12 (2025), 1441382.

[4] Dwain D Allan, Andrew J Vonasch, and Christoph Bartneck. 2022. Better than us: The role of implicit self-theories in determining perceived threat responses in HRI. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 215–224.

[5] Alexander Mois Aroyo, T Kyohei, Tora Koyama, Hideyuki Takahashi, Francesco Rea, Alessandra Sciutti, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Giulio Sandini. 2018. Will people morally crack under the authority of a famous wicked robot?. In *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 35–42.

[6] Wilma A. Bainbridge, Justin W. Hart, Elizabeth S. Kim, and Brian Scassellati. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3, 1 (2011), 41–52. doi:10.1007/s12369-010-0082-7

[7] Chester I Barnard. 1968. *The functions of the executive*. Vol. 11. Harvard university press.

[8] Sara C Benesh and Susan E Howell. 2001. Confidence in the courts: A comparison of users and non-users. *Behavioral Sciences & the Law* 19, 2 (2001), 199–214.

[9] Thomas Blass. 1999. The Milgram paradigm after 35 years: Some things we now know about obedience to authority. *Journal of Applied Social Psychology* 29, 5 (1999), 955–978.

[10] James P. Bliss, Shelby K. Long, and Nicole Karpinsky-Mosley. 2019. Cross-Cultural Reactions to Peacekeeping Robots Wielding Non-Lethal Weapons. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications, 2292–2297. doi:10.1177/1071181319631463

[11] Francis D Boateng. 2021. Police legitimacy in Africa: A multilevel multinational analysis. In *The rise of comparative policing*. Routledge, 109–124.

[12] Francis D Boateng and Zachary N Buckner. 2019. Police legitimacy in Asia: findings from a multilevel hierarchical non-linear analysis. *Policing and Society* 29, 7 (2019), 765–782.

[13] Matthew Bostrom. 2020. *Increasing police trust through normative alignment*. Ph. D. Dissertation. University of Oxford.

[14] Ben Bradford, Aziz Huq, Jonathan Jackson, and Benjamin Roberts. 2014. What price fairness when security is at stake? police legitimacy in South Africa. *Regulation & governance* 8, 2 (2014), 246–268.

[15] Anthony A Braga, Christopher Winship, Tom R Tyler, Jeffrey Fagan, and Tracey L Meares. 2014. The salience of social contextual factors in appraisals of police interactions with citizens: A randomized factorial experiment. *Journal of quantitative criminology* 30, 4 (2014), 599–627.

[16] Martim Brandao, Marina Jirotka, Helena Webb, and Paul Luff. 2020. Fair navigation planning: A resource for characterizing and designing fairness in mobile robots. *Artificial Intelligence* 282 (2020), 103259.

[17] Mai Lee Chang, Zachary Pope, Elaine Schaertl Short, and Andrea Lockerd Thomaz. 2020. Defining fairness in human-robot teams. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1251–1258.

[18] Mai Lee Chang, Greg Trafton, J Malcolm McCurry, and Andrea Lockerd Thomaz. 2021. Unfair! perceptions of fairness in human-robot teams. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 905–912.

[19] Yuan-Chia Chang, Daniel J Rea, Chi-Jung Lee, Takayuki Kanda, and Bing-Yu Chen. 2025. Designing Authoritative Presence in Social Robots. *ACM Transactions on Human-Robot Interaction* (2025).

[20] Vivienne Bihe Chi and Bertram F Malle. 2024. Interactive human-robot teaching recovers and builds trust, even with imperfect learners. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 127–136.

[21] Houston Claure. 2024. Designing for Fairness in Human-Robot Interactions. *arXiv preprint arXiv:2405.21044* (2024).

[22] Houston Claure, Inyoung Shin, J Gregory Trafton, and Marynel Vázquez. 2025. Did the Robot Really Intend to Harm Me? The Effect of Perceived Agency and Intention on Fairness Judgments. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 889–898.

[23] Jason A Colquitt. 2001. On the dimensionality of organizational justice: a construct validation of a measure. *Journal of applied psychology* 86, 3 (2001), 386.

[24] Martin Cooney, Masahiro Shiomi, Eduardo Kochenborger Duarte, and Alexey Vinel. 2023. A broad view on robot self-defense: Rapid scoping review and cultural comparison. *Robotics* 12, 2 (2023), 43.

[25] Derek Cormier, Gem Newman, Masayuki Nakane, James E Young, and Stephane Durocher. 2013. Would you do as a robot commands? An obedience study for human-robot interaction. In *International Conference on Human-Agent Interaction*, Vol. 13.

[26] Robert A Dahl. 1957. The concept of power. *Behavioral science* 2, 3 (1957), 201–215.

[27] John D'arcy and Tejaswini Herath. 2011. A review and analysis of deterrence theory in the IS security literature: making sense of the disparate findings. *European journal of information systems* 20, 6 (2011), 643–658.

[28] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[29] Dave DeNatale. 2023. *Crocker Park unveils new security guard: A robot named SAM.* Retrieved October 26, 2023 from https://www.wkyc.com/article/news/local/cuyahoga-county/crocker-park-new-security-guard-robot-sam/95-ecd06a56-ec46-4ce5-b309-87496e162ebd.

[30] Robin Shepard Engel. 2005. Citizens' perceptions of distributive and procedural injustice during traffic stops with police. *Journal of research in crime and delinquency* 42, 4 (2005), 445–481.

[31] Sibylle Enz, Martin Diruf, Caroline Spielhagen, Carsten Zoll, and Patricia A. Vargas. 2011. The social role of robots in the future—explorative measurement of hopes and fears. *Int. J. Soc. Robotics* 3 (2011), 263–271.

[32] Connor Esterwood, Ruijian Hannah Guan, Xin Ye, and Lionel P Robert. 2025. Virtually the Same or Realistically Different?: A Meta-analysis of Real vs.'Not So Real'Robots. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 559–568.

[33] Rocio De La Fe. 2024. *Three Security Robots Have Been Deployed into San Diego Neighborhoods.* https://www.cbs8.com/article/tech/autonomous-security-robots-deployed-tackle-crime-san-diego/509-aa0646e7-5d53-42fd-8247-659dbbba90c7 CBS8.

[34] Claes Fornell and David F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18, 1 (1981), 39–50.

[35] John RP French, Bertram Raven, and Dorwin Cartwright. 1959. The bases of social power. *Studies in social power* 150 (1959), 167.

[36] Darci Gallimore, Joseph B. Lyons, Thy Vo, Sean Mahoney, and Kevin T. Wynne. 2019. Trusting robocop: Gender-based effects on trust of an autonomous robot. *Front. Psychol.* 10 (2019), 482.

[37] Denise Geiskkovitch, Stela Seo, and James E Young. 2015. Autonomy, embodiment, and obedience to robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. 235–236.

[38] Denise Y Geiskkovitch, Derek Cormier, Stela H Seo, and James E Young. 2016. Please continue, we need more data: an exploration of obedience to robots. *Journal of Human-Robot Interaction* 5, 1 (2016), 82–99.

[39] OLGA GIL. 2023. Artificial Intelligence, Legitimacy and Control. (2023).

[40] Jerald Greenberg and Jason A Colquitt. 2013. *Handbook of organizational justice*. Psychology Press.

[41] Andrea Grundke, Jan-Philipp Stein, and Markus Appel. 2023. Improving evaluations of advanced robots by depicting them in harmful situations. *Computers in Human Behavior* 140 (2023), 107565.

[42] Christoffer Bjerre Haase, Margaret Bearman, John Brodersen, Klaus Hoeyer, and Torsten Risor. 2021. 'You should see a doctor', said the robot: Reflections on a digital diagnostic device in a pandemic age. *Scandinavian Journal of Public Health* 49, 1 (2021), 33–36.

[43] Joseph A Hamm, Rick Trinkner, and James D Carr. 2017. Fair process, trust, and cooperation: Moving toward an integrated framework of police legitimacy. *Criminal justice and behavior* 44, 9 (2017), 1183–1212.

[44] Kerstin S Haring, Ariana Mosley, Sarah Pruznick, Julie Fleming, Kelly Satterfield, Ewart J de Visser, Chad C Tossell, and Gregory Funke. 2019. Robot authority in human-machine teams: effects of human-like appearance on compliance. In *International conference on human-computer interaction*. Springer, 63–78.

[45] Kerstin S. Haring, Kelly M. Satterfield, Chad C. Tossell, Ewart J. De Visser, Joseph R. Lyons, Vincent F. Mancuso, Victor S. Finomore, and Gregory J. Funke. 2021. Robot authority in human-robot teaming: Effects of human-likeness and physical embodiment on compliance. *Frontiers in Psychology* 12 (2021), 625713. doi:10.3389/fpsyg.2021.625713

[46] Lyn Hinds and Kristina Murphy. 2007. Public satisfaction with police: Using procedural justice to improve police legitimacy. *Australian & New Zealand journal of criminology* 40, 1 (2007), 27–42.

[47] Mike Hough, Jonathan Jackson, and Ben Bradford. 2013. Legitimacy, Trust and Compliance: An Empirical Test of Procedural Justice Theory Using the European Social Survey. In *Legitimacy and Criminal Justice: An International Exploration*, Justice Tankebe and Alison Liebling (Eds.). Oxford University Press, Oxford. Available at SSRN: https://ssrn.com/abstract=2234339 or http://dx.doi.org/10.2139/ssrn.2234339.

[48] Ohad Inbar and Joachim Meyer. 2015. Manners matter: Trust in robotic peacekeepers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 59. Sage, Los Angeles, CA, 185–189.

[49] Ohad Inbar and Joachim Meyer. 2019. Politeness counts: Perceptions of peacekeeping robots. *IEEE Trans. Hum.-Mach. Syst.* 49, 3 (2019), 232–240.

[50] Jonathan Jackson. 2018. Norms, normativity, and the legitimacy of justice institutions: International perspectives. *Annual Review of Law and social science* 14, 1 (2018), 145–165.

[51] Jonathan Jackson, Ben Bradford, Mike Hough, Andy Myhill, Paul Quinton, and Tom R Tyler. 2012. Why do people comply with the law? Legitimacy and the influence of legal institutions. *British journal of criminology* 52, 6 (2012), 1051–1071.

[52] Jonathan Jackson, Ben Bradford, Betsy Stanko, and Katrin Hohl. 2012. *Just authority?: Trust in the police in England and Wales.* Willan.

[53] Elizabeth E Joh. 2017. Private security robots, artificial intelligence, and deadly force. *UCDL Rev.* 51 (2017), 569.

[54] Dan M. Kahan. 1999. The secret ambition of deterrence. *Harvard Law Review* 113 (1999), 414–500.

[55] Herbert C. Kelman. 1958. Compliance, Identification, and Internalization: Three Processes of Attitude Change. *Journal of Conflict Resolution* 2, 1 (1958), 51–60.

[56] Jennifer A. Kingson. 2023. Robots are your new office security guard. Axios. Retrieved October 26, 2023 from https://www.axios.com/2023/03/03/security-robots-artificial-intelligence

[57] Knightscope. 2023. Use cases: machine-as-a-Service. https://www.knightscope.com/who-we-serve. (2023). [Accessed March 20, 2025].

[58] Yung Hyeock Lee and Sujung Cho. 2020. The significance of instrumental pathways to legitimacy and public support for policing in South Korea: is the role of procedural fairness too small? *Crime, Law and Social Change* 73, 5 (2020), 575–603.

[59] Ela Liberman-Pincu, Amit David, Vardit Sarne-Fleischmann, Yael Edan, and Tal Oron-Gilad. 2021. Comply with Me: Using Design Manipulations to Affect Human–Robot Interaction in a COVID-19 Officer Robot Use Case. *Multimodal Technologies and Interaction* 5, 11 (2021), 71. doi:10.3390/mti5110071

[60] Jinchao Lin, April Rose Panganiban, Gerald Matthews, Katey Gibbins, Emily Ankeney, Carlie See, Rachel Bailey, and Michael Long. 2022. Trust in the danger zone: Individual differences in confidence in robot threat assessments. *Front. Psychol.* (2022), 1426.

[61] E Allan Lind and Tom R Tyler. 1988. *The social psychology of procedural justice.* Springer Science & Business Media.

[62] Beryl Lipton. 2024. *Drone as First Responder Programs Are Swarming Across the United States.* https://www.eff.org/deeplinks/2024/06/drone-first-responder-programs-are-latest-aerial-police-surveillance-push Electronic Frontier Foundation.

[63] Shelby K. Long, Nicole D. Karpinsky, and James P. Bliss. 2017. Trust of simulated robotic peacekeepers among resident and expatriate Americans. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61. Sage, Los Angeles, CA, 2091–2095.

[64] Alexander Lopez, Renato Paredes, Diego Quiroz, Gabriele Trovato, and Francisco Cuellar. 2017. Robotman: A security robot for human-robot interaction. In *2017 18th International Conference on Advanced Robotics (ICAR)*. IEEE, New York, NY, 7–12.

[65] Joseph B. Lyons, Sarah A. Jessup, and Thy Q. Vo. 2022. The Role of Decision Authority and Stated Social Intent as Predictors of Trust in Autonomous Robots. *Top. Cogn. Sci.* 16, 3 (Jan. 2022), 430–449.

[66] Joseph B. Lyons, Chang S. Nam, Sarah A. Jessup, Thy Q. Vo, and Kevin T. Wynne. 2020. The role of individual differences as predictors of trust in autonomous security robots. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. IEEE, New York, NY, 1–5.

[67] Joseph B. Lyons, Thy Vo, Kevin T. Wynne, Sean Mahoney, Chang S. Nam, and Darci Gallimore. 2021. Trusting autonomous security robots: The role of reliability and stated social intent. *Human factors* 63, 4 (2021), 603–618.

[68] Gabriela Marcu, Iris Lin, Brandon Williams, Lionel P. Robert Jr., and Florian Schaub. 2023. "Would I Feel More Secure With a Robot?": Understanding Perceptions of Security Robots in Public Spaces. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2 (2023), 1–34.

[69] Gerald Matthews, Jinchao Lin, April Rose Panganiban, and Michael D Long. 2019. Individual differences in trust in autonomous robots: Implications for transparency. *IEEE Transactions on Human-Machine Systems* 50, 3 (2019), 234–244.

[70] M. R. McGuire. 2021. The laughing policebot: Automation and the end of policing. *Polic. Soc.* 31, 1 (2021), 20–36.

[71] Kyle McLean. 2020. Revisiting the role of distributive justice in Tyler's legitimacy theory. *Journal of Experimental criminology* 16, 2 (2020), 335–346.

[72] Stanley Milgram. 2009. *Obedience to Authority: An Experimental View* (reprint ed.). Harper Perennial Modern Classics, New York. 256 pages.

[73] Kazuki Mizumaru, Satoru Satake, Takayuki Kanda, and Tetsuo Ono. 2019. Stop Doing It! Approaching Strategy for a Robot to Admonish Pedestrians. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 449–457. doi:10.1109/HRI.2019.8673133

[74] Eva Moravcová. 2016. Willingness to cooperate with the police in four central European countries. *European Journal on Criminal Policy and Research* 22, 1 (2016), 171–187.

[75] Roberta C Ramos Mota, Daniel J Rea, Anna Le Tran, James E Young, Ehud Sharlin, and Mario C Sousa. 2016. Playing the 'trust game'with robots: Social strategies and experiences. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 519–524.

[76] Andy Myhill and Paul Quinton. 2011. It'sa fair cop? Police legitimacy, public co-operation, and crime reduction: An interpretative evidence commentary. *National Policing Improvement Agency, Research, Analysis and Information, September 2011* (2011).

[77] Daniel S Nagin. 2013. Deterrence in the twenty-first century. *Crime and justice* 42, 1 (2013), 199–263.

[78] National Initiative for Building Community Trust and Justice. 2015. *Procedural Justice.* Technical Report. Office of Community Oriented Policing Services, Washington, DC.

[79] National Police Foundation. 2021. *21st Century Policing Task Force Report: The First Five Years.* Technical Report. National Police Foundation, Arlington, VA.

[80] Dietrich Oberwittler and Sebastian Roché. 2017. *Police-Citizen Relations Across the World: Comparing sources and contexts of trust and legitimacy.* Routledge.

[81] Nina Peršak. 2016. *Legitimacy and trust in criminal law, policy and justice: Norms, procedures, outcomes.* Routledge.

[82] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F Malle. 2018. What is human-like? decomposing robots' human-like appearance using the anthropomorphic robot (abot) database. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction.* 105–113.

[83] Alex R Piquero, Raymond Paternoster, Greg Pogarsky, and Thomas Loughran. 2011. Elaborating the individual difference component in deterrence theory. *Annual Review of Law and Social Science* 7, 1 (2011), 335–360.

[84] Princeton University Engineering. 2020. *AI tool gives doctors a new look at the lungs in treating COVID-19.* https://engineering.princeton.edu/news/2020/05/21/ai-tool-gives-doctors-new-look-lungs-treating-covid-19 Accessed: 2025-09-17.

[85] Alinnah Qiao. 2023. Under the Watchful, Unblinking Eye: Privacy Implications of the New York Police Department's Deployment of Autonomous Robots. *Harvard Journal of Law & Technology Digest* (2023). https://jolt.law.harvard.edu/digest/under-the-watchful-unblinking-eye-privacy-implications-of-the-new-york-police-departments-deployment-of-autonomous-robots

[86] Keely Quinlan. 2023. NYPD introduces two robot policing technologies, GPS dart system. https://statescoop.com/nypd-robot-policing-technologies-gps-speed-dart-system/. *StateScoop.* (2023).

[87] Paul Quinton. 2011. The impact of information about crime and policing on public perceptions: The results of a randomised controlled trial. *London: NPIA* (2011).

[88] Michael D Reisig, Jason Bratton, and Marc G Gertz. 2007. The construct validity and refinement of process-based policing measures. *Criminal justice and behavior* 34, 8 (2007), 1005–1028.

[89] Christian M. Ringle, Sven Wende, and Jan-Michael Becker. 2024. SmartPLS 4. https://www.smartpls.com/

[90] Lionel P Robert, Casey Pierce, Liz Marquis, Sangmi Kim, and Rasha Alahmad. 2020. Designing fair AI for managing employees in organizations: a review, critique, and design agenda. *Human–Computer Interaction* 35, 5-6 (2020), 545–575.

[91] Kantwon Rogers, De'Aira Bryant, and Ayanna Howard. 2020. Robot gendering: Influences on trust, occupational competency, and preference of robot over human. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, 1–7.

[92] Joshua Rust. 2018. Traditional action and traditional authority. *Max Weber Studies* 18, 2 (2018), 159–185.

[93] Roya Salehzadeh, Fareed Bordbar, Xiaoti Fan, Soroush Korivand, Glen Merritt, Darrin J Griffin, Christian Cousin, and Nader Jalili. 2025. Teleoperated Communication Robot: A Law Enforcement Perspective. *ACM Transactions on Human-Robot Interaction* 14, 2 (2025), 1–23.

[94] Shane P Saunderson and Goldie Nejat. 2021. Persuasive robots should avoid authority: The effects of formal and real authority on persuasion in human-robot interaction. *Science robotics* 6, 58 (2021), eabd5186.

[95] Catherine E Sembroski, Marlena R Fraune, and Selma Šabanović. 2017. He said, she said, it said: Effects of robot group membership and human authority on people's willingness to follow their instructions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 56–61.

[96] Herbert A Simon. 1951. A formal theory of the employment relationship. *Econometrica: Journal of the Econometric Society* (1951), 293–305.

[97] Starr J Solomon and Joselyne L Chenane. 2021. Testing the fair process heuristic in a traffic stop context: Evidence from a factorial study with video vignettes. *The British Journal of Criminology* 61, 4 (2021), 1086–1106.

[98] Ivan Y Sun, Yuning Wu, Rong Hu, and Ashley K Farmer. 2017. Procedural justice, legitimacy, and public cooperation with police: Does Western wisdom hold in China? *Journal of research in crime and delinquency* 54, 4 (2017), 454–478.

[99] Ivan Y Sun, Yuning Wu, Jianhong Liu, and Maarten Van Craen. 2019. Institutional procedural justice and street procedural justice in Chinese policing: The mediating role of moral alignment. *Australian & New Zealand Journal of Criminology* 52, 2 (2019), 272–290.

[100] Jason Sunshine and Tom R Tyler. 2003. The role of procedural justice and legitimacy in shaping public support for policing. *Law & society review* 37, 3 (2003), 513–547.

[101] Justice Tankebe. 2009. Public cooperation with the police in Ghana: Does procedural fairness matter? *Criminology* 47, 4 (2009), 1265–1293.

[102] The New York Times. 2023. NYPD Robot Gets Tryout to Patrol Times Square Subway. *The New York Times* (2023). Retrieved from https://www.nytimes.com/2023/09/22/nyregion/police-robot-times-square-nyc.html.

[103] The New York Times. 2024. NYPD Removes Robot from Times Square Subway. *The New York Times* (2024). Retrieved from https://www.nytimes.com/2024/02/02/nyregion/nypd-subway-robot-retires.html.

[104] Theodoros Theodoridis and Huosheng Hu. 2012. Toward intelligent security robots: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1219–1230.

[105] Masahiro Tsushima and Koichi Hamai. 2015. Public cooperation with the police in Japan: Testing the legitimacy model. *Journal of Contemporary Criminal Justice* 31, 2 (2015), 212–228.

[106] Tom Tyler and Steven Blader. 2013. *Cooperation in groups: Procedural justice, social identity, and behavioral engagement.* Routledge.

[107] Tom R Tyler. 1989. The psychology of procedural justice: A test of the group-value model. *Journal of personality and social psychology* 57, 5 (1989), 830.

[108] Tom R. Tyler. 1990. *Why People Obey the Law.* Yale University Press, New Haven, CT.

[109] Tom R Tyler. 1994. Psychological models of the justice motive: Antecedents of distributive and procedural justice. *Journal of personality and social psychology* 67, 5 (1994), 850.

[110] Tom R Tyler. 1997. The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Personality and social psychology review* 1, 4 (1997), 323–345.

[111] Tom R Tyler. 2003. Procedural justice, legitimacy, and the effective rule of law. *Crime and justice* 30 (2003), 283–357.

[112] Tom R Tyler. 2006. Psychological perspectives on legitimacy and legitimation. *Annu. Rev. Psychol.* 57, 1 (2006), 375–400.

[113] Tom R Tyler. 2006. *Why people obey the law.* Princeton university press.

[114] Tom R. Tyler. 2023. Whither legitimacy? Legal authority in the twenty-first century. *Annual Review of Law and Social Science* 19, 1 (2023), 1–17.

[115] Tom R Tyler and Jeffrey Fagan. 2008. Legitimacy and cooperation: Why do people help the police fight crime in their communities. *Ohio St. J. Crim. L.* 6 (2008), 231.

[116] Tom R Tyler and Yuen J Huo. 2002. *Trust in the law: Encouraging public cooperation with the police and courts.* Russell Sage Foundation.

[117] Tom R. Tyler, Jonathan Jackson, and Ben Bradford. 2013. Procedural Justice and Cooperation. In *Encyclopedia of Criminology and Criminal Justice*, Gerben Bruinsma and David Weisburd (Eds.). Springer, New York, 4011–4024. doi:10.1007/978-1-4614-5690-2_64

[118] Anna-Lisa Vollmer, Robin Read, Dries Trippas, and Tony Belpaeme. 2018. Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science robotics* 3, 21 (2018), eaat7111.

[119] Yvonne Wadsworth. 1998. What is Participatory Action Research? *Action Research International* Paper 2 (1998). Available online: http://www.scu.edu.au/schools/gcm/ar/ari/p-ywadsworth98.html.

[120] Max Weber. 2009. *The theory of social and economic organization.* Simon and Schuster.

[121] Scott E Wolfe, Justin Nix, Robert Kaminski, and Jeff Rojek. 2016. Is the effect of procedural justice on police legitimacy invariant? Testing the generality of procedural justice and competing antecedents of legitimacy. *Journal of quantitative criminology* 32, 2 (2016), 253–282.

[122] Kai Chi Yam, E-Yang Goh, Ryan Fehr, Randy Lee, Harold Soh, and Kurt Gray. 2022. When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human. *Journal of Experimental Social Psychology* 102 (2022), 104360.

[123] N'dea Yancey-Bragg. 2023. Police robots are on patrol. Now the questions about them are piling up. *USA TODAY* (2023). https://www.usatoday.com/story/news/nation/2023/10/31/nyc-police-robots-privacy-safety-concerns/71218199007/ Accessed: 2025-07-28.

[124] Xin Ye, Wonse Jo, Arsha Ali, Samia Cornelius Bhatti, Connor Esterwood, Hana Andargie Kassie, and Lionel Peter Robert. 2024. Autonomy acceptance model (aam): The role of autonomy and risk in security robot acceptance. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.* 840–849.

[125] Xin Ye and Lionel P Robert. 2023. Human security robot interaction and anthropomorphism: An examination of pepper, ramsee, and knightscope robots. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).* IEEE, 982–987.

[126] Xin Ye and Lionel P Robert. 2024. A human–security robot interaction literature review. *ACM Transactions on Human-Robot Interaction* 14, 2 (2024), 1–36.

[127] Xin Ye and Lionel P. Robert. 2025. Can Robots Take Over Security? A Review and Critique of Security Robots and Human Security Agents. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN).* IEEE.

[128] Xin Ye and Lionel P Robert. 2025. Security Robot Power and Acceptance: Exploring French and Raven's Five Forms of Power. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 1731–1736.

[129] Xin Ye and Lionel P Robert Jr. 2025. Rewarding Trust: How Reward Power Shapes Security Robot Acceptance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* SAGE Publications Sage CA: Los Angeles, CA, 10711813251357915.

[130] Sangseok You and Lionel P Robert Jr. 2018. Human-robot similarity and willingness to work with a robotic co-worker. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction.* 251–260.

[131] James Young and Derek Cormier. 2014. Can robots be managers, too. *Harvard Business Review* 4 (2014), 2014.

[132] Qiaoning Zhang, X Jessie Yang, and Lionel P Robert Jr. 2022. Individual differences and expectations of automated vehicles. *International Journal of Human–Computer Interaction* 38, 9 (2022), 825–836.

[133] Franklin E. Zimring. 2011. *The City That Became Safe: What New York Teaches About Urban Crime and Its Control.* Oxford University Press, New York.