Check for updates

Performance consequences of automated workplace control

Verena Dorner[1] and Gerlinde Fellner-Röhling[2]*

[1]Department of Information Systems and Operations Management, Institute of Digital Ecosystems, Vienna University of Economics and Business, Vienna, Austria, [2]Department of Economics, Vienna University of Economics and Business, Vienna, Austria

This study examines the impact of automated workplace control on employee performance and trust. In light of the rise in remote work and increasing use of algorithmic monitoring, we conduct a controlled experiment to investigate how workers' performance responds to control decisions made by an algorithm compared to a human. Moreover, we investigate spillovers on the subsequent trust in employers. Using a real-effort task and a trust game in an online experiment, we vary the source of control (human or algorithm) and control intensity (low, medium, high). Our findings reveal that control by a human principal (but not an algorithm) enhances worker performance, with no detrimental effects observed for higher control intensity. Despite the performance increase, being controlled by a human principal reduces trusting behavior on the extensive margin, i.e., the likelihood that agents entrust principals any positive amount of their endowment. Exploratory analyses suggest control by human principals is perceived more negatively than control by an algorithm. Our findings suggest that automated workplace control does not have negative ramifications for workers' performance, although it cannot generate the positive effects of human control either.

1 Introduction

Automated workplace control, algorithmic management, electronic performance monitoring and people analytics have come under much scrutiny in the last few years (Ball, 2021; Giermindl et al., 2022; Ravid et al., 2023). Advances in technology, such as wearable sensors and smartphones, ubiquitous computing, or algorithmic innovations, have expanded companies' interest in and ability to electronically monitor employees and automate managerial decisions (e.g., Edwards et al., 2018; Ravid et al., 2023). This trend has been further accelerated by the rise of platform-based gig work (Kadolkar et al., 2024) and a shift to remote office work, in particular since the COVID-19 pandemic (Ball, 2021; Kalischko and Riedl, 2021; Ravid et al., 2020).

While many employees appreciate the increased flexibility and reduced commuting costs of remote work, employers and contractors often fear that it increases opportunities for shirking. To make sure employees work the hours they are contracted to work, or perform their platform jobs as well as possible, many organizations increasingly adopt automated monitoring and control tools, like tracking clicks, taking screenshots (Ball, 2021; König, 2025), or assigning jobs or imposing penalties based on algorithms (Edwards et al., 2018; Kadolkar et al., 2024). Real-time monitoring is often a prerequisite for these systems (Kadolkar et al., 2024). Companies tout the advantages of such technologies,

like boosting productivity and enabling fairer performance evaluations. Critics, however, warn that they can erode trust between employers and employees, prompting workers to reduce their effort to the minimum required and potentially lowering overall productivity (Ball, 2021; Kalischko and Riedl, 2021; Giermindl et al., 2022). Our research addresses these competing claims. It is among the first to experimentally examine how automated control decisions affect worker performance and reciprocity toward employers in a controlled setting.

We focus on the area of algorithmic performance monitoring and control, which can be taken as one specific purpose (Ravid et al., 2020) under the broader term "electronic workplace surveillance," which refers to "any collection and processing of information, whether personally identifiable or not, for the purposes of influencing and managing those whose data have been garnered" (Ball, 2021; Lyon, 2001, p. 21).

Agency theory suggests that, in general, exerting control over workers, for instance by specifying minimum performance requirements or implementing audits, has a disciplining effect and can increase work effort in line with employers' interests. This is supported by findings from Dickinson and Villeval (2008), which indicate that workers respond positively to low and moderate levels of monitoring by the employer. However, the literature on motivation crowding-out (e.g., Frey, 1993) suggests that monitoring and controlling workers can backfire. Intrinsically motivated workers have been found to react most negatively to monitoring, reducing their performance to a greater extent than less motivated workers (Herz and Zihlmann, 2024). Similarly, high-cognitive ability workers have been shown to react more negatively to control in the shape of time and flexibility restrictions for task completion and scheduling (Rattini, 2023). The seminal paper by Falk and Kosfeld (2006) demonstrates that control exerted by a principal can incur hidden costs by signaling distrust to the agent. However, subsequent studies have found that the extent of the performance decline is not always sufficient to completely offset the incentive benefits of control (Ziegelmeyer et al., 2012). Additionally, the workplace setting influences outcomes: fewer workers reduce their effort under control in online environments compared to laboratory settings, while positive reciprocity in the absence of control is stronger in the lab (Schmelz and Ziegelmeyer, 2020). When monitoring is perceived as legitimate, hidden costs vanish (Schnedler and Vadovic, 2011). In situations with social ties, even weak monitoring reduces effort (Riener and Wiederhold, 2016). Workers have also been observed to adjust their behavior in unexpected ways, with performance improving in the monitored dimension (Nagin et al., 2002; Pierce et al., 2015), but deteriorating in other, non-monitored dimensions (Belot and Schröder, 2016; Herz and Zihlmann, 2024).

When it comes to automated workplace monitoring or control, performance effects seem to depend on the specific work setting. Three recent meta-analyses (König, 2025; Ravid et al., 2023; Siegel et al., 2022) do not find a consistent positive effect of automated workplace monitoring on performance, only on stress. However, the generalizability of these findings is limited due to methodological shortcomings, including a lack of longitudinal and multilevel field studies and the limited use of validated measurement scales. For instance, in her lab experiment, Strobel

(2025) found that employees choose significantly higher effort when a human manager sets a performance threshold for receiving a bonus compared to an automated process. However, the external validity of the findings may be limited by the fact that a one-shot interaction between a manager and a worker was implemented, and the worker did not engage in a real-effort task but stated a costly effort level instead.

Another important issue is the multiplicity of factors that moderate the effect of automated workplace monitoring and control on performance (Kalischko and Riedl, 2021; Kadolkar et al., 2024). Factors associated with higher performance are, for example, consequences and feedback mechanisms, while high task difficulty, and high intensity of monitoring are associated with lower performance (Kalischko and Riedl, 2021; Kadolkar et al., 2024). A recent online study mimicking remote work finds that low levels of control intensity increase performance, but raising the intensity further does not lead to a corresponding increase in performance—nor to a decrease (Mugerman et al., 2023).

Some studies hint at different reactions to workplace monitoring and control by men and women, but the evidence is limited and mixed. While a few studies report no significant differences in the acceptance of monitoring technologies (e.g., Jacobs et al., 2019; Zweig and Webster, 2003), others suggest that women tend to be more reticent and express greater concern about privacy (e.g., Stark et al., 2020). In contrast, one study finds the opposite pattern: women were more likely to take up monitored work when it offered higher remuneration and appeared less concerned about privacy issues (Liang et al., 2023).

Certainly, the rise of automated managerial decisions introduces a new layer to the employer-employee relationship: it reduces the direct (human) interaction between worker and manager and thereby limits social preferences, like reciprocity, to take effect (Giermindl et al., 2022). For example, when a principal waives control, it can signal trust and evoke positive reciprocal behavior by workers (Falk and Kosfeld, 2006). However, when the same decision is made by an algorithm on behalf of the principal, it may fail to convey that trust. In one experiment, workers exerted more effort when control was waived by a principal rather than by a neutral third party (Burdin et al., 2018). This raises an important question: Do workers respond to automated performance control (taken by an algorithm) in the same way as to human performance control (taken by a manager)?

Moreover, automation and the resulting reduction of social context in work settings raises broader concerns about its impact on the employer-employee relationship (Ball, 2021; Giermindl et al., 2022; Sallaz, 2015). Qualitative studies indicate that it can lead to contradictions between narratives of organizational care and coercive control (Sewell and Barker, 2006) as well as obscure hierarchies and reduced accountability (Giermindl et al., 2022; Sewell, 2021). In call centers, for example, a "regime of permanent pedagogy" has been observed to emerge, where continuous performance feedback and framing work as a learning process lead to high effort but also to internal resistance and low commitment to firm values (Sallaz, 2015). A survey of platform gig workers similarly found algorithmic management to be a double-edged sword: on the one hand, it reduced affective trust and reciprocity due to a perceived lack of care from the platform. On

the other hand, it enhanced trust by offering greater flexibility and by minimizing interference by human managers (Chen and Chen, 2025).

Taken together, the evidence on how automated monitoring and control affects workers' performance and trust remains mixed. Our study contributes to the hitherto fragmented literature by systematically examining three core questions: (i) whether workers respond differently to automated vs. human control, (ii) whether control intensity plays a role for the performance response to (automated or human) control and (iii) whether control has spillover effects on trust in the (human) manager. These spillover effects may in turn depend both on who exerts control and under which control intensity.

Our experiment models a workplace interaction in which a manager (principal) can exert control over a worker (agent). To reflect a realistic setting, the agent performs a real-effort task and the principal benefits from the agent's performance. Control over the agent can be exerted by setting a minimum performance threshold for the agent upfront. The agent receives payment for the task only if this threshold is met. Our first treatment variation concerns who makes the control decision: either the principal directly (*Human* treatment) or an algorithm acting on behalf of the principal (*Algorithm* treatment). The second treatment variation concerns the control intensity: we vary the minimum performance threshold that is applied in case of control in three steps. To assess the broader implications of (automated) control, we measure its effect on trust in a subsequent trust game (Berg et al., 1995), in which the agent takes the role of the trustor and the principal that of the trustee.

2 Materials and methods

2.1 Experimental design

Our experiment consists of 3 parts and a questionnaire. Subjects receive instructions for one part at a time and only know that other parts will follow but not what these parts are about. In the beginning they are informed that part 1 consists of one round and part 2 consists of five rounds. Participants earn points in all 3 parts, which convert to Euros at an exchange rate of 200 points = €1 for final payout. The experimental design and hypotheses are pre-registered (see https://osf.io/zpt28).

Part 1 serves to familiarize participants with the real-effort task, which requires using the mouse in order to place sliders exactly on a midpoint of an interval (Gill and Prowse, 2012). Participants can place up to 50 sliders that are displayed all on one screen and have a time limit of 90 s for the task. All participants work on the task in this part and earn a piece-rate wage of 12 points for every correctly placed slider.

In *Part 2*, participants are first randomly assigned to the roles of principal and agent and then random groups of one principal and one agent are formed. To avoid framing effects, we use an abstract description and refer to agents as players A and principals as players B. All participants are informed about the following procedure: part 2 has 5 rounds and in each round, agents will work on the same slider task as in part 1 for 90 s. However, before starting to work, a decision is taken whether the agent is subjected to control in the

particular round or not. Control is operationalized via a minimum performance threshold: agents have to place a minimum number of sliders correctly to earn the piece-rate wage of 12 points for every correctly placed slider. If they place fewer sliders correctly, they earn 0. In contrast, the principal always earns 12 points for each slider the agent places correctly.

To investigate the performance effects of automated control, we implement a 2×3 between-subjects design. The first treatment variable is the *source* of the control decision: the decision is taken by the principal (*Human* treatment) or by an algorithm on behalf of the principal (*Algorithm* treatment). The algorithm is a very simple one that chooses control or no control with equal probability. Participants were only informed that an algorithm will decide on whether to set a minimum performance threshold on behalf of the principal, but we gave no further information on how this algorithm will decide. In the *Human* treatment, control of the agent is the consequence of a deliberate decision of the principal. In the *Algorithm* treatment, the principal cannot affect the control choice. In either case, the consequence of control is the same for agents: to earn the piece-rate wage, their performance has to exceed the minimum performance threshold. Control is always costless for principals regardless of who takes the control decision.[1]

The second treatment variable is the *control intensity* that determines the agents' minimum performance requirement in case they are subjected to control. We vary control intensity in three steps: In treatment *low*, agents have to place at least 4 (out of 50) sliders correctly to earn the piece-rate wage, in treatment *medium* it is 13 sliders and in treatment *high* it is 26 sliders.[2]

Part 2 consists of five rounds of 90 s each and principal-agent pairs stay constant over rounds (partner matching). During the time agents perform the slider task, principals are engaged with an unrelated filler task.[3] After each round, before learning their earnings, agents are asked three questions on their subjective wellbeing in the work situation, in particular about how much effort

---

they subjectively exerted, how exhausted they got and how stressed they felt. The answer format for all 3 questions is a Likert scale ranging from 1 (very low / not at all) to 7 (very high / much).[4]

At the beginning of part 2, participants receive instructions about round 1 only. After completing round 1, they are informed that another 4 rounds will follow with the same procedure. Participants know that earnings from round 1 are always paid, and in addition, one of the subsequent 4 rounds is randomly selected for payout.

In *Part 3*, to investigate the consequences of (automated) control on the trust relation between principal and agent, participants take part in a standard one-shot trust game (Berg et al., 1995). The principal-agent pairing remains the same as in part 2. The agent is assigned the role of the trustor and the principal the role of the trustee. Agent and principal are endowed with 100 points each. The agent can decide to send any amount from 0 to 100 to the principal. The amount sent is tripled by the experimenter. After learning about the tripled amount, the principal decides whether and how many points to send back to the agent. Earnings from this part are paid out at the end of the experiment. The amount the agent sends to the principal indicates the level of trust in the principal.

In the *questionnaire*, participants fill in a demographic survey (age, gender) and answer several questions about their perception of a (no) control decision in their respective treatment. Agents state their work motivation in case of being controlled and not controlled. Principals are asked about the work motivation they expect the agent to have in these two cases. On the last screen, all participants are informed about their earnings in points from all parts of the experiment and their total payout in Euros.

The experiment was conducted online using O-tree (Chen et al., 2016) and participants were recruited from the subject pool of the experimental laboratory at the Vienna University of Economics and Business (WU labs). Recruitment was performed using ORSEE (Greiner, 2015), which ensures an unbiased assignment to treatments. In total, 496 subjects participated in our study (half of them agents, half of them principals). Four agent-principal pairs had to be excluded from the data analysis, since the agents completed the experiment on a touch screen device and ex-post reported technical difficulties with the slider task, leaving us with a final sample of 244 agents and principals, respectively. Table 1 gives an overview of the number of agents in each treatment. Seventy-nine participants were male, 161 were female and 4 preferred not to reveal their gender. Nearly all of the participants were students (62% Business or Economics, 10% Law, 12% Science or Engineering and 9% Humanities) and on average 23.7 years old. Sessions lasted ~ 40 min and the average payout was €9.1.

2.2 Research questions

To study the potential ramifications of automated control on workers' performance and reciprocity, we formulate four main research questions that we further break down into 10 testable

---

4 These questions are similar to the ones used by Dohmen and Falk (2011) to assess how hard participants worked in their experiment.

TABLE 1 Treatments and number of agents.

Source of control	Control intensity			
	Low (4)	Medium (13)	High (26)	Total
Human	37	34	44	115
Algorithm	43	52	34	129
Total	80	86	78	244

hypotheses based on the existing literature. Our first research question addresses the (ambiguous) findings in the literature on the general performance effects of managerial control.

Research question 1. Does control by a minimum performance requirement affect agents' effort?

While some studies find a positive incentive effect of monitoring agents' performance, as predicted by agency-theory (e.g., Dickinson and Villeval, 2008), the crowding-out hypothesis (Frey, 1993) suggests that such control can diminish agents' intrinsic motivation, potentially resulting in lower performance compared to a scenario without control. While control by monitoring agents and imposing sanctions increases their cost of shirking, our approach to managerial control involves requiring agents to meet a minimum performance level. This method reflects the type of control examined in the study of Falk and Kosfeld (2006), who find that it reduces workers' effort due to negative reciprocity. Consequently, in our first two hypotheses, we base our expectations regarding the performance effects of control on the findings of Falk and Kosfeld (2006). However, the innovative aspect of our study lies in investigating control within a more "realistic" setting, where agents engage in a real-effort task rather than merely reporting a costly effort level. Since we anticipate that control will be perceived as restricting agents' freedom to choose their performance, regardless of its source, we propose the same hypothesis for both the *Human* and *Algorithm* treatment.

HYPOTHESIS 1. *In the Human treatment, agents who are controlled by a minimum performance requirement show lower performance than agents who are not controlled.*

HYPOTHESIS 2. *In the Algorithm treatment, agents who are controlled by a minimum performance requirement show lower performance than agents who are not controlled.*

Our second research question examines the effects the severity of control by varying the minimum performance threshold that agents who are controlled have to meet.

Research question 2. Does the control intensity affect agents' performance?

As control restricts agents' freedom in their work task, implementing control under a higher minimum performance requirement is likely to intensify the effects of negative reciprocity. Therefore, we expect that higher control intensity will lead to lower performance among agents subjected to control compared to those under low control intensity. As we compare medium and high control intensity to low control intensity, and given that we have no

basis to expect this effect to differ between human and algorithmic control, we propose the following four hypotheses.

HYPOTHESIS 3a. *In the Human treatment, agents who are controlled show lower performance under medium intensity of control than under low intensity of control.*

HYPOTHESIS 3b. *In the Human treatment, agents who are controlled show lower performance under high intensity of control than under low intensity of control.*

HYPOTHESIS 4a. *In the Algorithm treatment, agents who are controlled show lower performance under medium intensity of control than under low intensity of control.*

HYPOTHESIS 4b. *In the Algorithm treatment, agents who are controlled show lower performance under high intensity of control than under low intensity of control.*

When agents are subjected to control, particularly in comparison to a situation without control, a higher level of control intensity conveys a stronger and more forceful signal. As a result, we expect performance differences between controlled and uncontrolled agents to be more pronounced when control intensity is higher. This forms the basis for our final two hypotheses related to the second research question.

HYPOTHESIS 5a. *Overall, the performance response to control is stronger in the medium control intensity treatment than in the low intensity treatment.*

HYPOTHESIS 5b. *Overall, the performance response to control is stronger in the high control intensity treatment than in the low intensity treatment.*

Our third research question addresses the key agenda of this study on the potential different reactions of agents to control by a human principal and by an automated system.

Research question 3. Does the source of control (*Human* vs. *Algorithm*) affect agents' effort?

According to motivation crowding theory (Frey and Jegen, 2001; Deci and Ryan, 1985), external interventions such as control can undermine agents' intrinsic motivation, either because they restrict autonomy or because they signal distrust and unkind intentions. In our setting, control in both the *Human* and *Algorithm* treatment involves a loss of autonomy, which may reduce agents' performance. However, in the *Human* treatment, control carries a stronger social signal: agents may perceive it as an intentional and potentially unkind act by the principal, thereby triggering negative reciprocity. In contrast, in the *Algorithm* treatment, control is less likely to be interpreted as a deliberate decision and thus does not convey the same sense of intentionality. As a result, while control may reduce performance in both treatments, we expect this effect to be more pronounced when control is imposed by a human. This aligns with evidence from neuroscience showing that reciprocal punishment of unfair behavior depends on the attribution of intentionality (Knoch et al., 2006). We thus state the following two hypotheses.

HYPOTHESIS 6. *Agents' peformance when being controlled is higher in the Algorithm treatment than the Human treatment.*

HYPOTHESIS 7. *The performance response to control is stronger in the Human treatment than in the Algorithm treatment.*

Our fourth and final research question addresses the potential spillover effects of control on the relationship between the principal and the agent, with a specific focus on trusting behavior in subsequent interactions.

Research question 4. Does being controlled affect trusting behavior?

Beyond its immediate effect on performance, control may have broader relational consequences. Control can signal distrust, which in turn may erode mutual trust. We expect that agents who were previously controlled, regardless of whether the control came from a human or an algorithm, will exhibit lower trust toward the principal than those who were not controlled. However, given that control by a human may more strongly signal distrust and unkind intent, we expect lower trust conditional on being controlled in the *Human* treatment. Furthermore, we expect the intensity of control to moderate this effect: control with higher minimum performance thresholds may be perceived as stronger signals of distrust or pressure, thereby reinforcing the negative spillover on trust. These considerations lead to the following four hypotheses.

HYPOTHESIS 8. *Agents who were controlled exhibit lower trusting behavior than agents who were not controlled.*

HYPOTHESIS 9. *Agents who were controlled exhibit lower trusting behavior in the Human treatment than in the Algorithm treatment.*

HYPOTHESIS 10a. *Agents who were controlled exhibit lower trusting behavior under medium control intensity than under low control intensity.*

HYPOTHESIS 10b. *Agents who were controlled exhibit lower trusting behavior under high control intensity than under low control intensity.*

3 Results

3.1 Descriptives on work performance and the use of control

We focus our analysis exclusively on the data of the 244 agents. Table 2 presents a descriptive overview of the average performance, measured by the number of correctly placed sliders separately for part 1 (without any control) and aggregated for all 5 rounds of part 2 of the experiment, separated by the two treatment variables *source of control* and *control intensity*.

In part 1, participants, on average, placed around 23 sliders out of the 50 possible correctly. This indicates that the high intensity of control (26 sliders) implemented in part 2 is binding for at least 50% of agents, while the low and medium intensities are not. However, this does not render the low and medium control intensities apriori irrelevant. Being controlled under any minimum performance requirement may lead to reactance such

TABLE 2 Average work performance (std. dev.) by treatments in part 1 and aggregated over all 5 rounds of part 2.

Source of control	Control intensity			
	Low (4)	Medium (13)	High (26)	Total
Part 1				
Human	19.76 (8.89)	21.88 (8.42)	23.48 (7.64)	21.81 (8.36)
Algorithm	23.40 (8.25)	23.25 (8.69)	24.21 (7.52)	23.55 (8.20)
Part 2 (aggregate)				
Human	27.00 (9.10)	28.85 (6.93)	28.45 (8.19)	28.11 (8.18)
Algorithm	28.73 (8.29)	28.67 (8.00)	30.55 (7.36)	29.19 (7.97)
Total	27.93 (8.71)	28.74 (7.59)	29.37 (7.90)	28.68 (8.08)

Note that in part 1, all agents work under the same conditions (no control) regardless of later treatment assignment.

that it may lower workers' motivation to exert effort beyond the threshold. Experimental evidence with stated effort (and effort cost) of workers has shown this effect (e.g., Falk and Kosfeld, 2006). As placing sliders is associated with real effort cost and being controlled potentially unfolds additional psychological cost, it is an empirical question whether workers respond negatively to control by exerting less effort above the threshold—even under a low (non-binding) minimum performance requirement.

In part 2, agents placed, on average, 28 out of the 50 possible sliders correctly. Only minor differences between treatments are visible at a first glance with a seemingly slightly positive reaction of performance to higher control intensity only in the *Algorithm* but not in the *Human* treatment.

Considering the actual use of control by principals, there is a small divergence between the *Algorithm* treatment (in which the actual frequency of control was 50.1% as expected by our implementation) and the *Human* treatment, in which control was used in 44.5% of the times. The likelihood of control stayed fairly constant over rounds (Periods 1 to 5, *Human*: 44.4%, 47.8%, 46.1%, 46.1%, 38.3%; *Algorithm*: 49.6%, 51.9%, 54.3%, 46.5%, 48.1%). Moreover, agents' exposure to control does not depend on the source of control since the control frequency does not significantly differ between the *Human* and *Algorithm* treatment in any of the rounds (two-sample test of proportions). Thus, human principals deliberately decide to control roughly as often as the algorithm randomly implements control on behalf of the principal. In addition, we checked whether the use of control differs across periods and the intensity levels of control. Results confirm that an agent's likelihood to be exposed to control does not depend on the source of control, the control intensity or the particular round.[5] Thus, any differential response of workers to control across different treatments, that we investigate in the following, is not driven by a different likelihood to be exposed to control.

In the subsequent analysis, we compare agents who are deliberately controlled with those deliberately not controlled in the *Human* treatment, and agents who are randomly controlled with those randomly not controlled in the *Algorithm* treatment.

In the latter, control is assigned exogenously by a random process, whereas in the former it is chosen endogenously by the principal. One of our core interests is whether a human principal's deliberate choice to control or not has different performance effects than the same decision made by an algorithm. Although the practical consequences for agents are identical—meeting a minimum performance requirement to earn the piece-rate wage—a human principal's choice can signal trust or distrust, whereas an algorithm's choice cannot. Accordingly, we compare the performance of agents who are controlled with those who are not in each round, distinguishing whether the decision originates from a human principal or an algorithm.

3.2 Work performance in round 1

As pre-registered, we first present the results from an OLS regression focusing on the work performance in round 1 of part 2 only. This is the first round where participants were divided into principals and agents and agents worked either under control or not.[6] The main dependent variable for our research questions 1 to 3 is the work performance of agents measured by the number of correctly placed sliders. The first main independent variable is the source of control that we capture by the indicator variable *Algorithm* that equals 1 for the algorithm decision on behalf of the principal and 0 for a human principal deciding whether to control or not. The second main independent variable is the control intensity that—for the sake of an easier interpretation—we capture with two dummy variables that equal 1 for *Medium intensity* (threshold of 13) and *High intensity* (threshold of 26), leaving *Low intensity* (threshold of 4) as the baseline. A further independent variable—that arises endogenously by the choice of the human principal or the algorithm—is *Control* that indicates whether the agent was subjected to control or not. In specification (1) of Table 3, we consider only the source of control (together with the information on whether the agent has been subjected to control or not), in specification (2), we consider the control intensity and in the final specification (3), we include all interaction effects of our main independent variables to obtain a complete picture of effects. In addition, we add sociodemographic controls with age and a dummy for women. Note that due to this inclusion, we loose 4 observations of agents who preferred not to reveal their gender.[7]

In neither specification can we identify a treatment effect of the source of control or the intensity of control on work performance. Agents seemingly show also no response to being controlled in any of the treatments. However, using only the first observation of work performance in part 2 of the experiment makes the estimations less efficient. To make better use of the repeated observations of work

---

5 Evidence can be found in the Supplementary material, Supplementary Table S1

---

6 Subjects are randomly assigned to treatments and roles (principals or agents), so that we do not expect work performance differences across treatments. However, we also used a difference-in-differences analysis on work performance in round 0 and 1, reproducing the same insights. We do not include this analysis here for the sake of parsimony in our results section.

7 All findings are robust against including these 4 observations by not considering sociodemographic controls.

TABLE 3 OLS regression of work performance in round 1.

Variable	(1)	(2)	(3)
Constant	28.021***	27.547***	27.076***
	(2.343)	(2.476)	(2.811)
Algorithm	0.959		1.205
	(1.234)		(2.630)
Control	-0.302	1.098	-0.233
	(1.689)	(1.895)	(2.973)
Control × Algorithm	0.812		2.548
	(2.137)		(3.752)
Medium intensity		1.954	2.257
		(1.642)	(2.387)
High intensity		2.460	2.443
		(1.608)	(2.373)
Control × Medium intensity		-1.954	-1.077
		(2.601)	(4.109)
Control × High intensity		-0.199	1.407
		(2.612)	(4.134)
Algorithm × Medium intensity			-0.538
			(3.306)
Algorithm × High intensity			0.352
			(3.231)
Algorithm × Control × Medium intensity			-2.063
			(5.297)
Algorithm × Control × High intensity			-3.217
			(5.238)
Female	-2.254**	-2.416**	-2.291**
	(1.107)	(1.113)	(1.130)
Age	0.001	-0.020	-0.031
	(0.086)	(0.083)	(0.084)
Observations	240	240	240
R-squared	0.027	0.036	0.050

Robust standard errors in parentheses.
** $p < 0.05$, *** $p < 0.01$.

performance, we exploit data on work performance over all rounds in the following.

3.3 Work performance throughout rounds

As it is likely that the effect of control and the exogenous treatment variations unfold over time, we make use of the repeated observations over rounds by re-estimating the OLS regression with data from rounds 1 to 5, including dummies for the rounds and clustering the standard errors at the individual level. We rely on the following estimation and present the results in Table 4 to test our

main hypotheses.

$$
\begin{aligned}
y_{it} = {} & \beta_1 + \beta_2 \text{Alg}_i + \beta_3 \text{Control}_{it} + \beta_4 (\text{Control}_{it} \times \text{Alg}_i) \\
& + \beta_5 \text{MediumIntensity}_i + \beta_6 \text{HighIntensity}_i \\
& + \beta_7 (\text{Control}_{it} \times \text{MediumIntensity}_i) \\
& + \beta_8 (\text{Control}_{it} \times \text{HighIntensity}_i) \\
& + \beta_9 (\text{Alg}_i \times \text{MediumIntensity}_i) \\
& + \beta_{10} (\text{Alg}_i \times \text{HighIntensity}_i) \\
& + \beta_{11} (\text{Alg}_i \times \text{Control}_{it} \times \text{MediumIntensity}_i) \\
& + \beta_{12} (\text{Alg}_i \times \text{Control}_{it} \times \text{HighIntensity}_i) \\
& + \mathbf{X}_i \gamma + \delta_t + \varepsilon_{it}
\end{aligned} \tag{1}
$$

The subscript i indexes individuals and t indexes time periods. The dependent variable y_{it}, which is the performance measure as number of correctly solved sliders, varies across individuals and rounds. Algorithm (Alg), Medium intensity, and High intensity are dummy variables that equal 1 for the respective treatments, leaving the Human and Low intensity treatment as the baseline. All treatment variables are assigned at the individual level and are time-invariant. Control equals 1 if the agent is controlled in round t and 0 otherwise. $\mathbf{X}_i$ indicates a vector of individual-level demographic control variables, which are age and a female dummy in our case. δ_t denotes round fixed effects. Standard errors are clustered at the individual level.

Our first research question concerns the agents' performance response to control by a minimum performance requirement. Specification (1) of Table 4 shows that control affects performance positively in the Human treatment (Control), which is opposite to our expectations formulated in Hypothesis 1.

Result 1. In the Human treatment, agents who are controlled show a higher performance than agents who are not controlled.

The negative (though only marginally) significant interaction effect Control × Algorithm indicates that the reaction to control is different in the Algorithm treatment. Indeed, the performance response to control in the Algorithm treatment is not significantly different from 0 ($\beta_3 + \beta_4 = 0$, $p = 0.70$). We thus find no support for Hypothesis 2.

Result 2. In the Algorithm treatment, agents' performance is not affected by control.

To account for the possibility that any (disciplining) effect of control may be stronger than one round, we re-estimated our main model of Table 4 by additionally considering a lagged control variable (lag of one period). The results, reported in Supplementary Table S2 in the Supplementary material, indicate that the control decision of the previous round does not significantly affect workers' performance in the current round, while all other findings in Table 4 are qualitatively unaffected.

The second research question concerns the reaction of agents to different control intensities. For Hypotheses 3a, 3b and 4a, 4b, we look at participants who are controlled in the Human and the Algorithm treatment, respectively, and compare their performance under medium and high control intensity to the baseline of low intensity. For that, we turn to specification (3) of Table 4. In the

TABLE 4 Pooled OLS regression of work performance with cluster-robust standard errors (rounds 1 to 5).

Variable		(1)	(2)	(3)
Constant	β_1	24.327***	24.923***	23.617***
		(2.165)	(2.284)	(2.673)
Algorithm	β_2	2.069*		2.352
		(1.073)		(2.411)
Control	β_3	2.355**	1.402	2.236
		(0.989)	(1.102)	(2.105)
Control × Algorithm	β_4	−2.149*		−1.666
		(1.127)		(2.286)
Medium intensity	β_5		1.449	2.543
			(1.415)	(2.342)
High intensity	β_6		0.624	0.202
			(1.477)	(2.393)
Control × Medium intensity	β_7		−1.746	−2.646
			(1.372)	(2.605)
Control × High intensity	β_8		1.720	2.645
			(1.449)	(2.519)
Algorithm × Medium intensity	β_9			−2.137
				(2.872)
Algorithm × High intensity	β_{10}			1.349
				(2.913)
Algorithm × Control × Medium intensity	β_{11}			1.733
				(2.897)
Algorithm × Control × High intensity	β_{12}			−2.308
				(2.859)
Female		−1.734*	−1.759*	−1.744*
		(0.985)	(0.991)	(0.987)
Age		0.097	0.086	0.089
		(0.070)	(0.071)	(0.070)
Round dummies			included	
Observations		1200	1200	1200
Clusters		240	240	240
R-squared		0.038	0.042	0.055

Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Human treatment, we observe no difference in performance under medium and low control intensity among those who are controlled ($\beta_5 + \beta_7 = 0$, $p = 0.95$). Similarly, there is no difference in performance across high and low control intensity ($\beta_6 + \beta_8 = 0$, $p = 0.14$). We thus find no support for Hypotheses 3a and 3b.

Result 3. In the *Human* treatment, the performance of agents who are controlled is not affected by the intensity of control.

We repeat this analysis for the *Algorithm* treatment and find similar results. Agents who are controlled do not perform differently under medium control intensity ($\beta_5 + \beta_7 + \beta_9 + \beta_{11} = 0$, $p = 0.75$) or under high control intensity ($\beta_6 + \beta_8 + \beta_{10} + \beta_{12} = 0$, $p = 0.29$) than under low intensity. Therefore, we find no support for Hypotheses 4a,b.

Result 4. In the *Algorithm* treatment, the performance of agents who are controlled is not affected by the intensity of control.

To test Hypotheses 5a and 5b we turn to specification (2) of Table 4. Agents' differential response to being controlled under medium compared to low intensity can be directly seen in the interaction effect Control × Medium intensity (β_7). Likewise, the interaction effect Control × High intensity (β_8) indicates whether agents react differently to control under high and low intensity. Both coefficients are not significantly different from zero, indicating that the different control intensities do not play a role for agents' reaction to being controlled.

Result 5. Agents' performance response to control is not affected by the control intensities.

Next, we investigate whether agents who are controlled perform differently in case the control is exerted by a human principal or an algorithm, independent of specific control intensities. To see this, we use specification (1) of Table 4. We cannot identify a performance difference between the *Human* and the *Algorithm* treatment for those who are controlled ($\beta_2 + \beta_4 = 0$, $p = 0.94$) and thus find no support for our Hypothesis 6.

Result 6. Agents' performance when being controlled does not depend on whether the control is exerted by a human principal or an algorithm.

Hypothesis 7 concerns the differential response to control by a human principal or an algorithm. In specification (1) of Table 4, we see that the interaction effect Control × Algorithm is marginally significant and negative (β_4, $p = 0.058$). This indicates that the observed positive response to control in the Human treatment is attenuated in the *Algorithm* treatment. In fact, as stated in Result 2, agents do not react positively to control by an algorithm. We therefore find weak support for Hypothesis 7 that the performance response to control is stronger in the *Human* than in the *Algorithm* treatment.

Result 7. Agents' (positive) performance response to control tends to be stronger when they are controlled by a human principal instead of an algorithm.

Remarkably, specification (1) of Table 4 shows a marginally significant and positive coefficient for the *Algorithm* treatment ($p = 0.055$). This indicates that agents who were *not* subjected to control show higher performance under an algorithm as principal than under a human principal. This is at odds with our initial expectation that human principals can signal trust and benevolence when deliberately abstaining from control and thus will be rewarded with higher performance of agents than an algorithm that cannot signal trust.

3.4 Behavior in the trust game

We turn now to the next set of pre-registered hypotheses to investigate the spillover effects of control decisions on subsequent interaction of the principal and the agent in the trust game. Agents, who were the trustors sent on average 56.41% of their endowment to the principal. In the following, we investigate whether agents' trusting behavior is affected by their experience of control, by the source of control and the control intensity. We analyze this question by looking at the intensive margin of trust, indicated by the amount sent, and at the extensive margin, indicated by the likelihood to send a positive amount instead of 0.

As the amount sent is censored from above (100) and below (0), Table 5 presents the results of Tobit regressions on the amount sent, i.e., the intensive margin of trust. The first three columns use the variable *Control* as the experience of a control decision in the first round of part 2, the second three columns alternatively use a definition of the variable *Control* as the number of times an agent has been controlled through all rounds of Part 2. On average, agents were controlled 2.4 times over the 5 rounds (*Human*: 2.2, *Algorithm*: 2.5). The results show that being controlled does not affect the amount agents send to principals in the trust game. Despite a negative coefficient, in neither specification is the variable Control significant. Moreover, the intensive margin of trust is not affected by the source of control [specification (2)] or the control intensity [specification (3)]. The control variable Female is significant and negative, indicating that women trust less than men.

As an additional test of how control affects trusting behavior, we estimate a probit regression of whether a positive amount was sent by the agent or not, i.e., we repeat the analysis on trusting behavior at the extensive margin. Table 6 shows the results.[8] The first three columns again consider *Control* as a dummy for being controlled in the first period, wheres the second three columns rely on a definition of *Control* by the number of times an agent was controlled over all 5 rounds.

The results of specifications (1) and (4) show that trust on the extensive margin is negatively affected by control, regardless of whether it is the initial control decision by the principal (algorithm) or the sum of all control decisions. This lends some support to our Hypothesis 8, yet only on the likelihood of trusting behavior, not the extent of trust.

Result 8. Control negatively affects the likelihood that agents trust the principal.

Specifications (2) and (5) show that agents who were controlled in the *Human* treatment do not differ in their likelihood to trust from agents who were controlled in the *Algorithm* treatment (Algorithm + Control × Algorithm=0, $p > 0.70$ for both specific.). We thus find no support for our Hypothesis 9.

Result 9. Agents exhibit similar trusting behavior regardless of whether they are controlled by a human principal or an algorithm.

For investigating the effect of the intensity of control on the trust behavior of controlled agents, we turn to specifications (3)

and (6). The likelihood to trust when being controlled is neither affected by medium or high control intensity, i.e., the sum of coefficients of medium or high control intensity and the respective interaction effect of Control with medium and high intensity are never significantly different from 0 in both specifications. We thus find no support for our Hypotheses 10a and 10b.

Result 10. Agents who are controlled exhibit similar trusting behavior under different control intensities.

Summarizing the results on our hypotheses, we find that agents—contrary to our initial expectations—react positively to control if control is exerted by a human principal. Control by an algorithm does not lead to a positive performance effect of control. For those agents who are controlled, however, performance does not differ under a human principal or an algorithm. The intensity of control, i.e., the minimum performance threshold the human principal or algorithm can set when controlling, does not affect work performance. Considering subsequent trust of agents after being controlled, we can conclude that the likelihood to trust decreases when being controlled (more often), but this does not depend on whether the control was exerted by a human principal or an algorithm. Again, different minimum performance thresholds do not affect trusting behavior.

3.5 Exploratory data analysis

In this section, we present exploratory data analysis that was not pre-registered due to the fact that no clear expectations from the literature could be identified. Nevertheless, these analyses can shed some more light on a potential heterogeneity of treatment effects (by gender) and the evaluation of control (in the different treatments), in addition to the reaction to it that we have analyzed so far.

3.5.1 Heterogeneous treatment effects by gender

We start by exploring whether men and women react differently to control by a human principal vs. an algorithm, both in the domain of work performance and trust. In Table 7 we present results from re-estimating the basic specification of our pooled OLS of work performance (see Table 4), separately for men and women.[9] The results show that all coefficients (except age) have the same sign for men and women, however, women tend to show stronger reactions. In particular, women show a strong and significant positive reaction to control in the *Human* treatment, while for men this positive reaction is not significant. Moreover, the marginally significant and negative interaction effect of Control and Algorithm indicates that women tend to respond less positively to control in the *Algorithm* than the *Human* treatment. Finally, women increase their performance over rounds while for men this is observable only for the last round. In sum, these findings align

---

8   The results are robust when using a linear probability model by an OLS regression.

9   We also re-estimated the other two specifications separating control intensities and the full model with all interactions. We found no evidence of heterogeneous treatment effects in these two specifications and therefore omit them for the sake of brevity.

TABLE 5 Tobit regressions of trust behavior (amount sent) on being controlled.

Control as:	Being controlled in first period			Number of times controlled		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	104.520***	106.492***	116.253***	107.924***	109.431***	106.147***
	(25.274)	(25.656)	(25.755)	(26.180)	(26.744)	(29.165)
Control	−6.711	−8.541	−24.599	−2.440	−2.675	−1.148
	(8.411)	(12.593)	(15.103)	(2.637)	(3.049)	(4.922)
Algorithm		−3.943			−4.462	
		(10.833)			(15.875)	
Control × Algorithm		3.595			1.036	
		(16.655)			(5.737)	
Medium intensity			−19.064			5.193
			(14.322)			(19.754)
High intensity			−18.517			−6.121
			(13.459)			(18.453)
Control × Medium intensity			27.628			−4.357
			(20.776)			(6.786)
Control × High intensity			23.675			0.160
			(19.845)			(6.098)
Female	−20.616**	−20.614**	−19.802**	−20.442**	−20.521**	−20.796**
	(9.764)	(9.784)	(9.695)	(9.802)	(9.731)	(9.745)
Age	−0.893	−0.891	−0.838	−0.931	−0.926	−0.830
	(0.999)	(0.998)	(0.982)	(0.998)	(0.998)	(1.011)
Observations	240					
Left-/right-/uncensored obs.	23/80/137					
Pseudo-R-squared	0.004	0.004	0.006	0.004	0.004	0.005

Dependent variable: amount sent by the agent in the trust game (0–100).

Robust standard errors in parentheses.

** $p < 0.05$, *** $p < 0.01$.

with literature that women react stronger to social stimuli than men and are more affected by the experimental conditions (Kamas and Preston, 2015).

Furthermore, we repeat the analysis of trusting behavior. In the estimations of Table 8, we focus on the main treatment variable source of control and on the principal's control decision in round 1, as considering the number of control decisions over all round does not add any further insights. Results show that men are less likely to trust the principal in the *Algorithm* than in the *Human* treatment when they have not been controlled. Being controlled by a human principal reduces mens' likelihood to trust (indicated by the negative coefficient of *Control*). The significant and positive interaction effect *Control × Algorithm* suggests that men are more likely to trust the principal if the control decision came from an algorithm than from a human. This suggests that using an algorithm to implement control decisions are less detrimental for men's trust than using control by a human principal. For women, trust decisions do not seem to depend on control or the source of control. The insights from this heterogeneity analysis can only serve as a first suggestion to where future studies may look more closely when it comes to potential gender differences. The limited number of observations of our study does not allow to draw unambiguous conclusions on gender differences in the reaction to algorithmic vs. human control.

3.5.2 Agents' wellbeing—subjective effort, exhaustion, and stress

In addition to studying work performance, for workplace design, it is vital to know whether agents' wellbeing is affected by being controlled and who exerts control. Therefore, after each round of working on the slider task, we asked agents to rate how much effort they subjectively have exerted, how exhausted they got, and how stressed they felt. The answer format to these three questions was a Likert scale ranging from 1 (not at all) to 7 (very much).

Table 9 gives a descriptive overview of the three different dimensions, aggregated over all five rounds and separated by the source of control. The mean evaluations after control and no

TABLE 6 Probit regressions of trust behavior (positive amount sent) on being controlled.

Control as:	Control in first period			Number of times controlled		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.224***	2.327***	2.798***	2.443***	2.403***	2.568***
	(0.573)	(0.598)	(0.760)	(0.619)	(0.629)	(0.708)
Control	−0.498**	−0.684*	−1.177**	−0.178**	−0.163*	−0.220*
	(0.245)	(0.379)	(0.522)	(0.075)	(0.087)	(0.127)
Algorithm		−0.184			0.141	
		(0.363)			(0.494)	
Control × Algorithm		0.338			−0.049	
		(0.485)			(0.160)	
Medium intensity			−0.774			0.020
			(0.531)			(0.586)
High intensity			−0.369			−0.342
			(0.571)			(0.547)
Control × Medium intensity			1.138*			−0.028
			(0.619)			(0.173)
Control × High intensity			0.637			0.158
			(0.686)			(0.174)
Female	0.474**	0.479**	0.519**	0.503**	0.502**	0.501**
	(0.237)	(0.232)	(0.229)	(0.237)	(0.231)	(0.223)
Age	−0.039*	−0.039*	−0.043**	−0.040**	−0.041**	−0.040**
	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)
Observations	240	240	240	240	240	240
Pseudo-R-squared	0.070	0.074	0.095	0.078	0.079	0.087

Dependent variable: 1 = agent sent a positive amount, 0 = otherwise.
Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

control hardly differ, indicating that being controlled does not play a role for the feeling of having exerted effort, being exhausted or stressed in our task. We further explore the effects that our treatment variations and control have on agents' wellbeing by regression analysis. Since all three variables are ordinal scaled, we use orderd logit regressions on the main treatment variables (*Algorithm*, *Medium* and *High intensity*), including the actual work performance in this round as a control, in addition to sociodemographic controls and round dummies. The results are displayed in Table 10. The main insights can be summarized as follows: actual work performance is positively correlated to subjective effort in the task, and agents feel more exhausted in the *Algorithm* than in the *Human* treatment. Being controlled by a minimum performance threshold does not affect agents' wellbeing.

3.5.3 Perception of control

In the post-experimental questionnaire, we asked agents for their view on the use or non-use of control in the specific work setting. In particular, we asked them to rate the decision of the principal (or algorithm) to control (not control) them along the following dimensions: unfair/fair, unkind/kind, malevolent/benevolent, and ungenerous/generous. Ratings were given on a slider scale and for each of the dimensions could range from 0 to 100. We can thus conduct within-subjects comparisons of how control is perceived compared to no control.

Table 11 gives a comprehensive overview of the mean perceptions of control and no control, separately for treatments *Human* and *Algorithm* and the three *control intensities*. The absence of control is significantly viewed more positively in all dimensions fairness, kindness, benevolence and generosity (Wilcoxon signed-rank tests: $p < 0.01$). The difference also holds when testing within each treatment group. This is remarkable since control does not decrease, but rather increases actual work performance (at least in the *Human* treatment), despite agents having a less positive view of it.

Moreover, aggregated over control intensity treatments (last two rows of Table 11), there is no difference in perception of the use of control between the treatments *Human* and *Algorithm*, with one exception. The use of control is perceived to be less benevolent

in the *Human* treatment than in the *Algorithm* treatment (51.3 vs. 58.0, Wilcoxon rank-sum test: $p = 0.04$). No control is perceived to be equally fair, kind and benevolent across the *Human* and *Algorithm* treatment.

Looking at the control intensity treatments separately, there are some notable differences in the perception of control and no control across the *Human* and *Algorithm* treatment. Under low control intensity, control by a human is perceived as less kind than control by an algorithm (54.8 vs. 71.2, $p = 0.03$), as less benevolent (50.9 vs. 70.1, $p = 0.03$) and also as less generous (53.8 vs. 67.6, $p = 0.03$). Under medium control intensity, the decision not to control is perceived as kinder (85.1 vs. 76.5, $p = 0.03$) and as more benevolent (79.1 vs. 68.0, $p = 0.03$) if it comes from a human principal rather than an algorithm.

Comparing the perception of control across different control intensities, aggregated over the *Human* and *Algorithm* treatment, reveals that the use of control is perceived less positive in all dimensions the higher the intensity of control (Kruskal Wallis tests, $p < 0.01$). This contrasts with the finding that work performance is not affected by the control intensities, regardless of whether control is exerted or not.

In sum, we find some suggestive evidence that the decision to use control is viewed less positive when it comes from a human principal rather than an algorithm and, vice versa, the decision not to control is viewed as more positive when a human makes it. This aligns with our initial expectations that human principals can give social signals by deciding to control or not, which an algorithm can not. However, this effect operates in favor of involving automation in control decisions: according to our explorative findings, an automated control decision has less detrimental effects on agents' perceptions than a human control decision. Yet, this effect is sensitive to the control intensity in our study, which requires more thorough investigation in future studies.

3.5.4 Work motivation under control and no control

In the final part of our explorative data analysis, we address agents' subjective motivation to work under control or no control. Agents could state their work motivation for the case they are controlled and for the case they are not, on a five point Likert scale

TABLE 7 Heterogeneous treatment effects on work performance.

Variable	Men	Women
Constant	28.298***	20.024***
	(2.973)	(2.800)
Algorithm	2.304	2.146
	(1.770)	(1.372)
Control	1.817	2.730**
	(1.594)	(1.226)
Control × Algorithm	−1.388	−2.674*
	(1.879)	(1.386)
Age	−0.053	0.193*
	(0.104)	(0.099)
Round 2	0.290	1.552***
	(0.608)	(0.396)
Round 3	1.044	1.974***
	(0.747)	(0.394)
Round 4	1.311	2.663***
	(0.795)	(0.447)
Round 5	2.550***	2.483***
	(0.836)	(0.527)
Observations	395	805
R-squared	0.030	0.037

Pooled OLS regression, clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 8 Heterogeneous treatment effects on the likelihood of trust.

Variable	Men	Women
Constant	6.841***	2.035***
	(1.093)	(0.742)
Control	−5.228***	-0.206
	(0.443)	(0.411)
Algorithm	−4.802***	0.519
	(0.353)	(0.521)
Control × Algorithm	4.973***	-0.331
	(0.606)	(0.655)
Age	−0.042	−0.024
	(0.037)	(0.026)
Observations	79	161
Pseudo R-squared	0.167	0.034

Probit regression, robust standard errors in parentheses.
Dependent variable: 1 = agent sent a positive amount, 0 = otherwise.
*** $p < 0.01$.

TABLE 9 Descriptive overview of subjective work experience aggregated over all rounds.

	Effort		Exhaustion		Stress	
Group	Human	Algorithm	Human	Algorithm	Human	Algorithm
No control	5.72 (1.60)	5.85 (1.55)	4.30 (1.90)	4.40 (1.96)	4.74 (1.80)	4.82 (1.75)
Control	5.91 (1.51)	5.94 (1.48)	4.35 (2.12)	4.31 (2.11)	4.86 (1.98)	4.89 (1.96)
Total	5.82 (1.54)	5.88 (1.53)	4.33 (2.01)	4.35 (2.03)	4.80 (1.88)	4.85 (1.88)

(1=very low motivation, 5=very high). Table 12 summarizes the mean values (and standard deviations), separated by treatments.

Overall, agents state that their work motivation when being controlled is lower than when not controlled (3.6 vs. 3.9, Wilcoxon signed rank test: $p < 0.01$). Separating the analysis by the source of control reveals that only in the *Human* treatment do agents state lower motivation under control (3.39 vs. 4.01, $p < 0.01$), not in the *Algorithm* treatment (3.81 vs. 3.83, $p = 0.86$). This pattern is consistent across all control intensities. Moreover, stated work motivation when being controlled is higher in the *Algorithm* than in the *Human* treatment (3.81 vs. 3.39, Wilcoxon rank sum test: $p = 0.03$), while work motivation when not controlled does not differ between treatments.

In line with the findings on agents' wellbeing and perception of control, this suggests that automated control is less harmful than human control. Remarkably, the lower work motivation under control by a human principal does not translate to agents' actual performance. On the contrary, in the *Human* treatment, control even increases performance. Yet, that agents do not express lower motivation under control in the *Algorithm* treatment aligns with Result 2 that agents' performance is not affected by control if it is exerted by an algorithm.

TABLE 10 Ordered logit regressions of subjective work experience.

Variable	(1)	(2)	(3)
	Effort	Exhaustion	Stress
Control	0.185	−0.015	0.179
	(0.152)	(0.154)	(0.152)
Algorithm	−0.135	0.484**	0.292
	(0.218)	(0.230)	(0.217)
Medium intensity	−0.009	−0.160	−0.107
	(0.272)	(0.272)	(0.260)
High intensity	−0.143	0.316	0.333
	(0.265)	(0.276)	(0.281)
Work performance	0.049***	0.009	0.011
	(0.015)	(0.012)	(0.013)
Female	0.322	0.074	0.381*
	(0.222)	(0.218)	(0.212)
Age	0.002	0.023	0.024
	(0.019)	(0.023)	(0.023)
Period 2	0.182**	0.308***	0.033
	(0.087)	(0.065)	(0.078)
Period 3	0.046	0.477***	0.021
	(0.104)	(0.093)	(0.101)
Period 4	0.134	0.593***	0.016
	(0.122)	(0.110)	(0.108)
Period 5	0.223*	0.704***	0.061
	(0.120)	(0.111)	(0.112)
Observations	1200	1200	1200

Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4 Discussion

This study aims to investigate how automated workplace control affects workers' performance and trust. In our experiment, agents have to perform a real-effort task that is paid by a piece-rate and benefits a human principal. In the treatment with human control, the principal can decide to exert control over the agent by setting a minimum performance threshold that the agent has to pass to earn the wage. In the treatment with automated control,

TABLE 12 Agents' stated motivation under control and no control by treatment.

Control intensity	Human		Algorithm	
	Control	No control	Control	No control
Low	3.30 (1.08)	3.97 (0.73)	3.86 (0.97)	3.77 (1.00)
Medium	3.47 (1.21)	4.09 (0.93)	3.83 (1.02)	3.87 (0.97)
High	3.41 (1.11)	3.98 (0.93)	3.71 (1.03)	3.85 (0.93)
Total	3.39 (1.12)	4.01 (0.86)	3.81 (1.00)	3.83 (0.96)

Mean values (standard deviations).

TABLE 11 Mean perceptions of control and no control use.

Intensity		Fair		Kind		Benevolent		Generous	
		Hum.	Alg.	Hum.	Alg.	Hum.	Alg.	Hum.	Alg.
Low	Control	63.0	74.5	54.8	71.2	50.9	70.1	53.8	67.6
	No control	77.1	83.0	77.5	82.1	71.0	77.2	75.3	80.1
Medium	Control	63.1	63.8	51.1	55.2	50.6	54.7	50.9	57.4
	No control	86.7	79.0	85.1	76.5	79.1	68.0	83.5	75.1
High	Control	57.1	46.9	50.0	38.1	52.2	47.6	48.2	41.7
	No control	79.4	81.5	79.5	77.2	72.4	72.9	74.4	75.8
Total	Control	60.8	62.9	51.9	56.0	51.3	58.0	50.8	56.6
	No control	80.8	81.0	80.5	78.6	73.9	72.3	77.4	77.0

an algorithm implements a random control decision on behalf of the principal. In addition to the source of control, we vary the control intensity by three different levels of minimum performance thresholds between-subjects. Subsequent to the effort task, agents interact with the same principal in a trust game, where the agent is the trustor and the principal the trustee. We complement the analysis of behavioral data with questions on the perception of control and work motivation.

Our findings reveal various subtle patterns that enhance our understanding of algorithmic management and its behavioral effects. We find that control by a human principal significantly increases agents' work performance, while control by an algorithm does not. This contrasts with the notion that control inherently reduces intrinsic motivation (Frey, 1993; Falk and Kosfeld, 2006) and instead aligns more closely with agency theory predictions (Dickinson and Villeval, 2008). The positive performance response to human-imposed control suggests that, in our context, control was not perceived as a signal of distrust strong enough to crowd out motivation. Notably, the source of control (human vs. algorithm) does not moderate the performance of workers who are actually subjected to control. However, performance gains from control appear to be driven primarily by the human treatment, indicating that algorithmic control lacks the same motivational impact, possibly due to the absence of perceived intentionality or social interaction. Contrary to common concerns about the rigidity of algorithmic systems, we observe no detrimental effects of higher control intensity on performance in either treatment. This suggests that raising performance thresholds, when applied transparently and uniformly, does not necessarily impair effort.

When it comes to trust, our findings suggest that being subjected to control reduces the likelihood of trusting behavior in a subsequent interaction, irrespective of whether control was exerted by a human or an algorithm. However, among those who do choose to trust, the amount sent does not differ based on previous control experience. Thus, control primarily impacts the extensive margin of trust, i.e., whether trust is initiated, rather than the intensive margin, i.e., how much is entrusted. This points to a subtle but important consequence: control may erode the formation of the trust relationship without necessarily affecting its depth once formed.

Exploratory results suggest that agents perceive human control decisions more negatively than algorithmic ones. Control imposed by a human was associated with lower self-reported motivation and was rated as less fair, kind, and benevolent. Interestingly, these subjective impressions do not align with actual performance: while motivation is lower under human control, performance is higher. This disconnect raises important questions about the role of perceived fairness and emotional response in driving behavior under workplace control. We also observe some indications of gender differences. Female agents respond more positively in terms of performance to human control, while male agents exhibit reduced trust when controlled by a human rather than an algorithm. Although exploratory, these patterns align with prior evidence that men and women may differ in their social sensitivity (Kamas and Preston, 2015).

Our results add valuable insights to the ongoing discussions about algorithmic monitoring in the workplace. While algorithmic control avoids some of the negative affective responses that human control elicits, it also appears less effective at motivating effort. The erosion of trust following control, regardless of its source, indicates that any form of enforced performance can carry social costs, even when productivity is maintained or enhanced. For organizations, this highlights a trade-off between enforcing control to increase output and maintaining trust and motivation among employees. On a positive side for automating managerial decisions, our findings indicate that algorithmic control is not more harmful than human control in terms of performance or trust. However, they underscore that algorithmic systems may not unfold the same motivational effects as human control. In contrast, empirical studies of algorithmic monitoring and control in real-world settings have found substantial negative effects of algorithmic control, attributing them to reduced autonomy, opaque algorithmic decision-making and employees' perceptions of being reduced to data points and consequently dehumanized (Giermindl et al., 2022).

This study relies on a stylized experimental setting with a simple real-effort task and short-term interactions. While this allows for causal identification, it limits generalizability to complex, long-term employment relationships. The algorithm in our study was deliberately designed to be opaque and random, but the monitoring and control decision was quite simple and transparent, without the possibility of large adverse effects like the real-world equivalent of losing the job or facing reputational penalties that go beyond the loss of performance-based pay for one round. Real-world algorithms often differ in perceived legitimacy, transparency, and learning capacity, which may alter behavioral responses. Future research should investigate how these and other algorithmic design features interact with performance and trust outcomes. Some features like time pressure likely increase the perception of control (Kocher and Sutter, 2006). Gamification features like leaderboards, intended to enhance motivation, effort and learning, introduce competition, which can lead to performance increases but reduced engagement for certain groups of employees (e.g., Michinov and Michinov, 2025; Nebel et al., 2017). Longitudinal experiments could assess how repeated exposure to control—human or algorithmic—shapes trust and productivity as well as perceptions of organizational care and resistance behaviors (Sallaz, 2015; Sewell and Barker, 2006; Yam et al., 2022) and social status (Jago et al., 2024) over time. The suggestive gender differences we observe also warrant further exploration, potentially in a more diverse labor context.

Overall, in our experimental setting, automated control does not negatively impact performance, but it also does not enhance it to the extent that human control does. While control, irrespective of its source, tends to weaken trust, its effect on performance depends on who implements it. These results contribute to the expanding body of literature on the social and behavioral implications of algorithmic systems in the workplace.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. Generative AI was used for text editing only.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frbhe.2025.1647057/full#supplementary-material

References

Ball, K. (2021). *Electronic Monitoring and Surveillance in the Workplace. Literature Review and Policy Recommendations*. Publications Office of the European Union: Luxembourg.

Belot, M., and Schröder, M. (2016). The spillover effects of monitoring: a field experiment. *Manag. Sci*. 62, 37–45. doi: 10.1287/mnsc.2014.2089

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games Econ. Behav*. 10, 122–142. doi: 10.1006/game.1995.1027

Burdin, G., Halliday, S., and Landini, F. (2018). The hidden benefits of abstaining from control. *J. Econ. Behav. Organ*. 147, 1–12. doi: 10.1016/j.jebo.2017.12.018

Chen, D. L., Schonger, M., and Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *J. Behav. Exp. Finance*. 9, 88–97. doi: 10.1016/j.jbef.2015.12.001

Chen, J., and Chen, M. (2025). Humanized algorithms matter! From enabling social exchange factors to building affective trust mechanism for facilitating gig workers' task performance. *Pers. Rev*. 54, 636–659. doi: 10.1108/PR-07-2024-0688

Deci, E. L., and Ryan, R. M. (1985). *Intrinsic Motivation and Self-determination in Human Behavior*. Plenum Press: New York. doi: 10.1007/978-1-4899-2271-7

Dickinson, D., and Villeval, M.-C. (2008). Does monitoring decrease work effort? the complementarity between agency and crowding-out theories. *Games Econ. Behav*. 63, 56–76. doi: 10.1016/j.geb.2007.08.004

Dohmen, T., and Falk, A. (2011). Performance pay and multidimensional sorting: productivity, preferences, and gender. *Am. Econ. Rev*. 101, 556–590. doi: 10.1257/aer.101.2.556

Edwards, L., Martin, L., and Henderson, T. (2018). *Employee Surveillance: The Road to Surveillance is Paved with Good Intentions*. Available online at: https://ssrn.com/abstract=3234382 (Accessed August 18, 2018).

Falk, A., and Kosfeld, M. (2006). The hidden costs of control. *Am. Econ. Rev*. 96, 1611–1630. doi: 10.1257/aer.96.5.1611

Frey, B. (1993). Does monitoring increase work effort? The rivalry between trust and loyalty. *Econ. Inq*. 31, 663–670. doi: 10.1111/j.1465-7295.1993.tb00897.x

Frey, B. S., and Jegen, R. (2001). Motivation crowding theory. *J. Econ. Surv*. 15, 589–611. doi: 10.1111/1467-6419.00150

Georganas, S., Tonin, M., and Vlassopoulos, M. (2015). Peer pressure and productivity: the role of observing and being observed. *J. Econ. Behav. Organ*. 117, 223–232. doi: 10.1016/j.jebo.2015.06.014

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., and Redzepi, A. (2022). The dark sides of people analytics: reviewing the perils for organisations and employees. *Eur. J. Inf. Syst*. 31, 410–435. doi: 10.1080/0960085X.2021.1927213

Gill, D., and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *Am. Econ. Rev*. 102, 469–503. doi: 10.1257/aer.102.1.469

Greif-Winzrieth, A., Dorner, V., Könemann, J., and Fellner-Röhling, G. (2025). "The heart of effort: Revealing heart rate patterns in real-effort tasks," in *Lecture Notes in Information Systems and Organization* (Cham: Springer Nature), 287–298. doi: 10.1007/978-3-031-71385-9_25

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *J. Econ. Sci. Assoc.* 1, 114–125. doi: 10.1007/s40881-015-0004-4

Herz, H., and Zihlmann, C. (2024). Adverse effects of control? Evidence from a field experiment. *Exp. Econ.* 27, 469–488. doi: 10.1007/s10683-024-09823-3

Jacobs, J. V., Hettinger, L. J., Huang, Y.-H., Jeffries, S., Lesch, M. F., et al. (2019). Employee acceptance of wearable technology in the workplace. *Appl. Ergon.* 78, 148–156. doi: 10.1016/j.apergo.2019.03.003

Jago, A. S., Raveendhran, R., Fast, N., and Gratch, J. (2024). Algorithmic management diminishes status: an unintended consequence of using machines to perform social roles. *J. Exp. Soc. Psychol.* 110:104553. doi: 10.1016/j.jesp.2023.104553

Kadolkar, I., Kepes, S., and Subramony, M. (2024). Algorithmic management in the gig economy: a systematic review and research integration. *J. Organ. Behav.* 46, 1–24. doi: 10.1002/job.2831

Kalischko, T., and Riedl, R. (2021). Electronic performance monitoring in the digital workplace: conceptualization, review of effects and moderators, and future research opportunities. *Front. Psychol.* 12:633031. doi: 10.3389/fpsyg.2021.633031

Kamas, L., and Preston, A. (2015). Can social preferences explain gender differences in economic behavior? *J. Econ. Behav. Organ.* 116, 525–539. doi: 10.1016/j.jebo.2015.05.017

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Kocher, M. G., and Sutter, M. (2006). Time is money—time pressure, incentives, and the quality of decision-making. *J. Econ. Behav. Organ.* 61, 375–392. doi: 10.1016/j.jebo.2004.11.013

König, C. J. (2025). Electronic monitoring at work. *Annu. Rev. Organ. Psychol. Organ. Behav.* 12, 321–342. doi: 10.1146/annurev-orgpsych-110622-060758

Liang, C., Peng, J., Hong, Y., and Gu, B. (2023). The hidden costs and benefits of monitoring in the gig economy. *Inf. Syst. Res.* 34, 297–318. doi: 10.1287/isre.2022.1130

Lyon, D. (2001). *Surveillance Society*. Maidenhead: McGraw-Hill Education (UK).

Michinov, N., and Michinov, E. (2025). More competition, less interaction: gamifying lectures using a leaderboard reduces female students' social engagement. *J. Comput. High. Educ.* 1–27. doi: 10.1007/s12528-025-09438-4

Mugerman, Y., Winter, E., and Yafeh, T. (2023). *Keep an eye on your employees, but not two: An experimental study of online monitoring, performance, and effort.* Researchgate. Available online at: https://www.researchgate.net/profile/Yevgeny-Mugerman/publication/369197418_Keep_an_Eye_on_Your_Employees_but_Not_Two_An_Experimental_Study_of_Online_Monitoring_Performance_and_Effort/links/640f8def66f8522c38a0464e/Keep-an-Eye-on-Your-Employees-but-Not-Two-An-Experimental-Study-of-Online-Monitoring-Performance-and-Effort.pdf (Accessed March 13, 2023).

Nagin, D. S., Rebitzer, J. B., Sanders, S., and Taylor, L. J. (2002). Monitoring, motivation, and management: the determinants of opportunistic behavior in a field experiment. *Am. Econ. Rev.* 92, 850–873. doi: 10.1257/00028280260344498

Nebel, S., Schneider, S., Beege, M., and Rey, G. D. (2017). Leaderboards within educational videogames: the impact of difficulty, effort and gameplay. *Comput. Educ.* 113, 28–41. doi: 10.1016/j.compedu.2017.05.011

Pierce, L., Snow, D. C., and McAfee, A. (2015). Cleaning house: the impact of information technology monitoring on employee theft and productivity. *Manag. Sci.* 61, 2299–2319. doi: 10.1287/mnsc.2014.2103

Rattini, V. (2023). Worker autonomy and performance: evidence from a real-effort experiment. *J. Econ. Manag. Strateg.* 32, 300–327. doi: 10.1111/jems.12511

Ravid, D. M., Tomczak, D. L., White, J. C., and Behrend, T. S. (2020). Epm 20/20: a review, framework, and research agenda for electronic performance monitoring. *J. Manag.* 46, 100–126. doi: 10.1177/0149206319869435

Ravid, D. M., White, J. C., Tomczak, D. L., Miles, A. F., and Behrend, T. S. (2023). A meta-analysis of the effects of electronic performance monitoring on work outcomes. *Pers. Psychol.* 76, 5–40. doi: 10.1111/peps.12514

Riener, G., and Wiederhold, S. (2016). Team building and hidden costs of control. *J. Econ. Behav. Organ.* 123, 1–18. doi: 10.1016/j.jebo.2015.12.008

Sallaz, J. J. (2015). Permanent pedagogy: how post-fordist firms generate effort but not consent. *Work Occup.* 42, 3–34. doi: 10.1177/0730888414551207

Schmelz, K., and Ziegelmeyer, A. (2020). Reactions to (the absence of) control and workplace arrangements: experimental evidence from the internet and the laboratory. *Exp. Econ.* 23, 933–960. doi: 10.1007/s10683-020-09666-8

Schnedler, W., and Vadovic, R. (2011). Legitimacy of control. *J. Econ. Manag. Strateg.* 20, 985–1009. doi: 10.1111/j.1530-9134.2011.00315.x

Sewell, G. (2021). Algorithmic management and the reorganization of work. *Organ. Stud.* 42, 685–701.

Sewell, G., and Barker, J. R. (2006). Coercion versus care: using irony to make sense of organizational surveillance. *Acad. Manag. Rev.* 31, 934–961. doi: 10.5465/amr.2006.22527466

Siegel, R., König, C. J., and Lazar, V. (2022). The impact of electronic monitoring on employees' job satisfaction, stress, performance, and counterproductive work behavior: a meta-analysis. *Comput. Hum. Behav. Rep.* 8:100227. doi: 10.1016/j.chbr.2022.100227

Stark, L., Stanhaus, A., and Anthony, D. L. (2020). "i don't want someone to watch me while I'm working:" gendered views of facial recognition technology in workplace surveillance. *J. Assoc. Inf. Sci. Technol.* 71, 1074–1088. doi: 10.1002/asi.24342

Strobel, C. (2025). The impact of process automation on performance. *J. Behav. Exp. Econ.* 117:102377. doi: 10.1016/j.socec.2025.102377

Yam, K. C., Goh, E.-Y., Fehr, R., Lee, R., Soh, H., and Gray, K. (2022). When your boss is a robot: workers are more spiteful to robot supervisors that seem more human. *J. Exp. Soc. Psychol.* 102:104360. doi: 10.1016/j.jesp.2022.104360

Ziegelmeyer, A., Schmelz, K., and Ploner, M. (2012). Hidden costs of control: four repetitions and an extension. *Exp. Econ.* 15, 323–340. doi: 10.1007/s10683-011-9302-8

Zweig, D., and Webster, J. (2003). Personality as a moderator of monitoring acceptance. *Comput. Hum. Behav.* 19, 479–493. doi: 10.1016/S0747-5632(02)00075-4