Working Without Wages: The Consequences of Widespread Pay Delays*

Dauda Musa
The Spartak Consult

Daniel J. Sonnenstuhl (JMP)
University of Chicago

November 16, 2025
Frequently updated, view latest version here

Abstract

We show that firms in low-income countries frequently withhold employee wages and study how workers respond to this widespread practice. Using original survey data from Lagos, Nigeria, we document that 30% of workers across firms of all sizes report delayed or unpaid salaries. To examine how workers respond to wage withholding, we conduct a field experiment in which we establish a firm in Nigeria, reach over 1,700 jobseekers through our recruitment process, and hire 600 for multi-month employment. Unpaid wages increase employees' initial effort, as outstanding balances raise the amount workers expect to receive in the future, without affecting absenteeism or total hours worked. The prevalence of wage withholding creates uncertainty that induces worker selection. Credibly signaling salary reliability increases job take-up by 25%, attracting workers who would otherwise be unwilling to accept wage employment but are no more productive. Combining intensive- and extensive margin estimates suggests that, in our setting, firms lose at most 0.2% in productivity from engaging in wage withholding. This gives firms little incentive to refrain from the practice, while workers are willing to forgo more than 30 percent of the monthly minimum wage for reliable pay.

*Daniel Sonnenstuhl thanks James Robinson, Eduardo Montero, and Greg Lane for their continuous guidance and support throughout this project and the Ph.D. We also thank Nathan Barker, Chris Blattmann, Christina Brown, Martin Castillo Quintana, Emanuele Colonnelli, Josh Dean, Oeindrila Dube, Steven Durlauf, Francisco Eslava, Yana Gallen, Josh Gottlieb, Amen Jalal, Ed Jee, Anne Karing, Erin Kelley, Michael Kremer, Francesco Loiacono, Danielle Nemschoff, Stephen O'Connell, Evans Osabuohien, Raúl Sánchez de la Sierra, Ben Shaver, Shaoda Wang, Chris Woodruff and audiences at the University of Chicago, SIOE, RES PhD Conference, CSAE, DevPEC, AFE Conference, STEG Early Career Workshop and RWI Berlin for helpful comments and discussions. We thank the entire field team for their outstanding work throughout this project. This research is funded by The Weiss Fund for Research in Development Economics, the Development Economics Center and the Becker Friedman Institute, all at the University of Chicago. This document is an output from the research initiative 'Structural Transformation and Economic Growth' (STEG), a programme funded by the Foreign, Commonwealth & Development Office (FCDO), contract reference STEG_LOA_3650_Sonnenstuhl. The views expressed are not necessarily those of FCDO. This research has been approved by the University of Chicago Social and Behavioral Sciences IRB (Protocol: IRB24-0989) and the Covenant Health Research Ethics Committee (Protocol: CHREC/529/2024) at Covenant University. This experiment was pre-registered at the AEA registry under AEA registry number: AEARCTR-0015132.

1 Introduction

The organization of modern economies rests on a simple exchange: workers supply labor to firms and are paid wages in return. Yet in many low-income countries, firms frequently disregard this exchange by withholding employees' wages. Stories of unpaid salaries — from nurses in Ghana (News Ghana, 2025b) to dry-cleaners in Zimbabwe (Gausi, 2018) — regularly appear in newspapers across Sub-Saharan Africa.[1] In Nigeria, the setting of this study, we document the scale of the problem and estimate that 30 percent of employees have experienced substantially delayed or unpaid salaries.[2]

This uncovers a practice that not only harms millions of affected workers but also contrasts sharply with economic theory. The usual conception of work entails reliable wage payments, typically underpinned by employees' ability to seek legal redress when firms fail to pay. But even if enforcement of labor contracts through legal channels is difficult, as in many low-income countries, workers still have avenues to hold employers accountable — for example, by retaliating in response to wage disputes (e.g., Bewley, 1998; Krueger and Mas, 2004; Mas, 2006). Moreover, by avoiding unreliable employers, high-ability workers can deprive firms of productive labor, making wage withholding even less attractive. These mechanisms should give employers strong incentives to pay on time. Yet, as we document, wage withholding is widespread, raising a central question: how do workers in low-income countries actually respond to unpaid wages?

In this paper, we study how workers respond to both delayed wages and the payment risk created by widespread wage withholding. We implement a randomized controlled trial (RCT) that varies the reliability of wage payments and addresses two empirical challenges. First, we link verified instances of delayed pay — otherwise rarely observed — to granular worker

[1] Examples of unpaid wages span multiple countries and sectors, including the sports, oil, and medical industries in Nigeria (Fernandes-Brough, 2023; Akuopha, 2023; Punch, 2023; News Ghana, 2025a), sanitation work in Côte d'Ivoire (Africanews, 2024), garment manufacturing in Senegal (IndustriALL Global Union, 2025), journalism in Cameroon (Animbom et al., 2022), professional football in Gabon (BBC Sport, 2019), civil service in Sudan (Radio Tamazuj, 2023), and mining in Zambia (Jeffay, 2024), among others.

[2] Authors' calculations from original survey data; Section 2.2 discusses these numbers in more detail.

performance measures, enabling us to study on-the-job responses. Second, we recruit individuals not actively seeking wage employment in the current environment and randomly guarantee on-time pay, allowing us to observe job acceptance under counterfactual payment reliability — essential for studying selection. We complement the experiment with a descriptive survey documenting the prevalence of wage withholding and a theoretical framework to understand the trade-offs workers face when their wages remain unpaid.

The experiment yields three main results: (i) delaying wages modestly *increases* employee effort rather than eliciting retaliatory behavior — an effect driven by the prospect of future pay; (ii) signaling salary reliability raises job take-up, especially among individuals otherwise disinterested in wage employment but (iii) does not affect workforce quality or composition. These findings suggest that, in a context with slack labor markets, firms face minimal productivity losses from engaging in wage withholding, helping explain its prevalence.

We begin our analysis by documenting the prevalence and economic significance of wage arrears in Nigeria. Drawing on original survey data ($n = 1,279$), we find that 30 percent of workers across firms of all sizes report experiencing delayed or unpaid wages. Workers rarely take action in response — only 19 percent left their job, and only one percent pursued recovery through legal channels. These experiences translate into concerns about employers' reliability: 47 percent of respondents worry that small employers, and 31 percent that large employers, fail to pay wages as agreed.

After documenting the prevalence of wage arrears, we analyze how workers respond to this practice. To understand the trade-offs workers face when their wages remain unpaid, we develop a theoretical framework in which worker reactions to delayed payments play a central role. We model wages as non-binding promises. Unpaid amounts accumulate as outstanding balances that roll over into subsequent periods and increase the amount a worker *ought to* be paid in the future. This framework generates three predictions. First, the relationship between wage arrears and optimal worker effort is ambiguous and depends on the institutional environment. In settings such as Nigeria, where payment uncertainty is high, unpaid wages can initially incentivize workers to exert greater effort to increase the likelihood of continued employment and eventual repayment. Second, when uncertainty about wage payment is high, increasing payment certainty raises workers' willingness to accept wage employment. Third, the effect of selection on workforce composition is ambiguous: if driven by productivity, new workers will be more productive; if driven by risk preferences, they will be more risk averse.

To empirically study workers' behavior in response to wage withholding, we conducted a field experiment in Lagos. We incorporated a local firm and recruited individuals for in-person image-labeling tasks tailored to this experiment to precisely measure productivity. We recruited two distinct samples. The first sample consisted of active jobseekers who proactively responded to physical job advertisements ($n = 638$), allowing us to study behavioral responses among typical employees. The second sample, reached through in-person recruitment, consisted of individuals not seeking wage employment ($n = 1,079$) and enabled us to study how salary uncertainty might deter people from entering wage employment.[3] All recruitees were randomly assigned to one of three employment conditions that varied in salary certainty: arm 1, a control condition with no information about payment reliability; arm 2, a low-certainty condition explicitly stating a 10-50 percent chance of nonpayment in a given pay cycle; and arm 3, a high-certainty condition guaranteeing timely wage payments.

This design allows us to identify three effects that directly map to the predictions of our conceptual framework: (a) selection into wage employment, (b) employees' behavioral responses to delayed pay, and (c) workforce composition effects. All recruitees were required to signal their willingness to work under their assigned terms by attending an in-person orientation day.[4] Drawing on the in-person-recruited sample, this setup allows us (a) to study selection into wage employment in response to higher salary certainty. We then hired 600

[3] All recruitment was conducted in collaboration with a local agency whose recruiters presented themselves as employees of a third-party company hiring on behalf of a client.

[4] Each recruitee received a letter outlining the terms of potential employment. While these terms were fixed, participants understood that actual hiring depended on vacancy availability and confirmed interest.

recruitees who attended the orientation day across two employment rounds. The first round, comprising only job-advertisement recruitees, is used (b) to identify employees' behavioral responses to delayed pay. Among those whose terms explicitly mentioned possible payment delays, we randomly assigned salary delays during employment to study employees' ensuing behavior.[5] The second round, which included both recruitment samples, is used (c) to examine workforce composition. By comparing the productivity and characteristics of workers newly attracted by reliable pay to those of typical jobseekers, we assess whether higher salary certainty alters workforce composition.

First, we find that being owed salary leads to a modest but statistically significant increase in employee productivity: image classification performance increased by 0.5 percent ($p < 0.01$) on our continuous accuracy index, corresponding to a 5.6 percent ($p < 0.05$) improvement in flawless task completions. These effects are robust across a range of checks and driven by employees with weaker outside options, for whom job loss or forgoing unpaid salaries would be especially costly. Results also remain unchanged for the ex-ante designated "high-stakes" images, where employees were asked to exert greater effort but also had an opportunity to retaliate against the employer. We find no meaningful changes in absenteeism and hours worked in response to salary delays. Overall, these findings suggest that in environments with limited legal recourse, unpaid wages can motivate greater worker effort rather than provoke retaliation, consistent with the first prediction of our conceptual framework.

Second, we find that salary uncertainty deters individuals with higher outside options from selecting into wage employment. Receiving a terms-of-employment letter explicitly conveying higher salary certainty significantly increased job take-up, especially among in-person-recruited individuals who had initially expressed no interest in the job. Take-up in this group increased by 11 percentage points ($p < 0.01$), corresponding to a 25 percent effect size. In contrast, job acceptance exceeded 95 percent among individuals who actively responded to job advertisements and have measurably lower outside options, with no significant variation

[5] Ethical considerations are discussed in Section 4.7.

across treatment arms — consistent with our model's prediction that salary certainty should matter primarily for those with stronger outside options. We interpret these findings as evidence that widespread salary uncertainty discourages participation in wage employment.

Third, we find that selection into employment induced by higher salary certainty is not driven by differences in worker productivity but rather reflects variation in risk preferences. We first examine whether workers attracted by reliable pay differ in productivity or characteristics from typical employees. To do so, we use a standard LATE framework (Imbens and Angrist, 1994), treating assignment to the high-certainty employment condition as an instrumental variable for job take-up. Identifying the characteristics of compliers — individuals induced to accept the job by the high-certainty employment condition — shows that they are statistically indistinguishable from always-takers on observable characteristics, employees who accept wage employment even in the control condition. However, there is suggestive evidence that compliers are more risk-averse than always-takers and more closely resemble never-takers. Second, we assess productivity differences between the in-person recruited sample and the job-advertisement recruited sample and find that they perform equally well. These results suggest that while reliable payment attracts additional workers, it does not substantially affect workforce quality or composition. This implies limited productivity consequences for firms in our setting, given the slack labor market.[6]

As a final exercise, we leverage randomized wage offers to quantify individuals' monetary valuation of salary certainty. We calculate the marginal rate of substitution between salary certainty and the wage offer to obtain a willingness-to-pay (WTP) measure.[7] Individuals place substantial value on salary certainty, with an estimated WTP of about 25,000 Nigerian

[6]The latest official *labor underutilization* rate, reported for 2024, was 14.5% (National Bureau of Statistics, 2024). This measure combines unemployment and time-related underemployment, i.e., individuals who work less than full-time and would like to work more. In a context where many people hold informal jobs that are rarely full-time, this is the most appropriate indicator for assessing labor market slack. As an empirical reference for labor market slackness, we were able to recruit more than 300 employees within roughly three weeks of posting job advertisements for our experiment.

[7]Strictly speaking, this measure captures the amount of salary individuals were willing to forgo for greater salary certainty, but interpreting it as a WTP remains conceptually valid.

Naira (NGN) — approximately USD 15 — *per month*, corresponding to over 100 percent of the median weekly wage. To validate these estimates, we use a choice experiment administered during the job interview or information session and obtain very similar results.

Our findings suggest that environments with weak enforcement and limited worker recourse create firm-side moral hazard. Firms can engage in wage withholding while facing limited risk of legal challenge and incurring only minor productivity losses. To gauge the productivity implications for firms, we combine the lower bound of our intensive-margin estimate with the upper bound of our extensive-margin estimate in a back-of-the-envelope calculation. Even under this deliberately unfavorable interpretation for firms, the resulting productivity loss is at most about 0.2 percent in our setting with slack labor markets.[8] Losses would likely be larger in environments where labor markets are tighter and replacing workers is more costly. At the same time, wage withholding imposes substantial welfare costs on workers, reflected in their high valuation of reliable pay. In a second back-of-the-envelope calculation, we show that this implies wage withholding is a Pareto-inefficient practice. Firms could obtain similar effort responses to those produced by wage withholding at lower cost, while not reducing worker welfare, by modestly lowering baseline wages, guaranteeing timely payment, and offering small performance bonuses.[9]

One caveat to our findings is that we do not capture potential general-equilibrium effects on firms. In our data, workers who avoid wage employment are more likely to enter self-employment, where they compete with firms. This competition may impose additional costs on firms — an important avenue for future research. A further limitation is that, although wage withholding directly reduces workers' welfare, it may also generate offsetting effects that we cannot measure. Firms might, for instance, be less willing to hire if withholding wages were not an option.

We make four contributions. First, prior to this study, systematic evidence on wage arrears in

[8] See Appendix H.1 for details.
[9] See Appendix H.2 for details.

low-income countries was limited to insulated public-sector settings (Flynn and Pessoa, 2014; Buehren et al., 2018). More broadly, documentation of wage arrears and analysis through a macroeconomic lens exist for post-Soviet Russia (Alfandari and Schaffer, 1996; Clarke, 1998; Lehmann et al., 1999; Earle and Sabirianova, 2002; Gerry et al., 2004; Lehmann and Wadsworth, 2007; Earle and Peter, 2009) and Ukraine (Boyarchuk et al., 2005).[10] However, to the best of our knowledge, no prior research has systematically documented the prevalence of wage arrears in low-income countries.

Second, we connect wage arrears to the literature on deferred compensation (e.g., Becker and Stigler, 1974; Lazear, 1979; Huck et al., 2011) and optimal wage setting (e.g., Shapiro and Stiglitz, 1984; Akerlof and Yellen, 1990; Bewley, 1999). Our conceptual framework illustrates how salary delays can increase the continuation value of the employment relationship by raising the amount a worker expects to receive in the future.

Third, our analysis builds on existing research that examines worker performance and behavior, especially in low-income countries (e.g. Falk, 2007; Kaur et al., 2015; Breza et al., 2018; Freeman et al., 2025; Kaur et al., 2025). We demonstrate, theoretically and empirically, that salary delays can serve as an incentive mechanism by motivating workers to increase effort. This finding aligns closely with evidence showing that performance-linked incentives increase worker productivity in low-income contexts (Bandiera et al., 2007; Ashraf et al., 2014, 2018; Guiteras and Jack, 2018) and more generally (e.g. Lazear, 2000).

Fourth, we also contribute to the literature on labor-market frictions in low-income countries (see Breza and Kaur (2025) for a recent overview). A growing body of evidence documents worker reluctance toward wage employment, driven by preferences for flexibility (Blattman and Dercon, 2018) — potentially linked to habit formation (Cefala et al., 2024) — complementary labor supply (Donald and Grosset, 2024), redistributive pressures (Carranza et

[10]While wage arrears are typically not an issue in industrialized economies, related phenomena such as wage theft — especially minimum-wage violations — are documented in low-skilled labor markets in these contexts (Bernhardt et al., 2009; Milkman et al., 2010; Robinson et al., 2011; Galvin, 2016; Clemens and Strain, 2022).

al., 2022), and cultural norms (Cassan et al., 2021; Oh, 2023). Kapoor (2025) investigates how liquidity constraints and limited firm commitment shape contractual preferences among job-seeking day laborers in Indian spot-labor markets. We go beyond this evidence to show that salary uncertainty — arising broadly from widespread wage arrears — constitutes a general and significant deterrent to wage employment. More broadly, our findings highlight that payment uncertainty shapes individual behavior, consistent with evidence from other contexts (e.g. Dunn, Gottlieb, Shapiro, Sonnenstuhl and Tebaldi, 2024).

2 Salary Uncertainty in Nigeria: Context and Empirical Patterns

In this section, we briefly outline the broader context of our study and present novel descriptive evidence on salary uncertainty and wage withholding in urban Nigeria. We document six new facts that describe its scale, incidence, and workers' responses. The institutional and economic context is essential to conceptualize both the emergence and consequences of widespread wage arrears. Although our data and contextual description are specific to Nigeria, the setting shares key features with labor markets across many low-income countries.

2.1 Context

The setting for this study is Lagos, a megacity in southwestern Nigeria with approximately 18 million inhabitants. Lagos is Nigeria's largest city and economic hub, although Nigeria remains among the poorest countries in the world, with a GDP per capita of $1,596.9 in 2023 (World Bank, 2023 USD). Nigeria's economy is heavily dependent on oil, making it sensitive to fluctuations in global oil prices and causing recurrent inflationary pressures. This volatility exacerbates financial insecurity for firms and workers alike. While the global technology boom has also reached Nigeria, making Lagos home to one of Africa's most vibrant startup and tech sectors, the majority of people still work in more traditional economic activities. Retail and manufacturing, for instance, are important sectors both in Lagos and across Nigeria (PwC, 2024). Many of these businesses are notably small and clustered in market-like settings; as an example, Lagos's *Computer Village* is a market named after its concentration of small-scale IT and phone retailers. Most of these shops are independently owned, typically employing only a few individuals. In fact, micro, small, and medium-sized enterprises collectively employ around 80 percent of Nigeria's workforce and generate over 40 percent of its GDP (PwC, 2020, 2024).[11] Micro enterprises constitute over 95 percent of these businesses, making the median enterprise in Nigeria effectively a one-person operation. The vast majority of these enterprises operate informally, with limited access to formal financial systems and minimal regulatory oversight.

Lack of effective governance and regulatory enforcement also characterizes Nigeria's broader institutional environment. A recent U.S. Department of State Human Rights Report describes widespread corruption, substantial shortcomings in enforcing employment laws — violations of minimum wage regulations and basic labor standards are rarely investigated — and an overall grim human rights situation in Nigeria (US Department of State, Bureau of Democracy, Human Rights and Labor, 2023). This characterization resembles scholarly assessments of Nigeria's rule-of-law environment, which document compromised judicial independence, political interference, and routine disregard for judicial decisions (e.g. Akomolafe, 2021; Igwe, 2021). Nigeria's consistently low global position in the World Justice Project's rule-of-law index reflects these conditions.[12] Such conditions are not unique to Nigeria; unreliable law enforcement and the lack of an impartial, well-functioning judiciary are common institutional characteristics in low-income countries (Sánchez de la Sierra, 2021; Sánchez de la Sierra et al., 2024).

2.2 Survey Evidence: Six New Facts on Wage Withholding

We conducted an original survey to document new empirical facts about wage withholding in

[11]Micro enterprises generally have 10 or fewer employees, small enterprises 11–50 employees, and medium-sized enterprises 51–200 employees.

[12]The World Justice Project is an independent organization that publishes annual global rule-of-law rankings based on factors such as constraints on government power, absence of corruption, regulatory enforcement, and protection of fundamental rights. Nigeria has consistently ranked 120th or lower out of 142 countries over the past three years.

Nigeria. Between June and August 2025, we surveyed 1,279 individuals in Lagos. To obtain a quasi-representative sample of the city's low- and medium-skilled labor force, enumerators approached respondents at busy public locations using a randomized skip pattern. Because our focus is on workers' experiences with wage withholding, we restricted the sample to individuals who reported current or previous employment. The main survey findings are presented in Figure 1, which summarizes six new facts about wage withholding.

Fact 1: Wage withholding is prevalent. Panel (a) shows that salary delays and non-payments are widespread. Overall, 29.9 percent of survey respondents reported having experienced at least one form of salary difficulty. We define salary difficulty as experiencing one or more of the following: (i) delayed salary payments, (ii) partial salary payments, or (iii) complete non-payment of wages. Among respondents, seven percent reported receiving only partial payments and four percent reported no payment at all.[13] The most common problem, however, was delayed salary payments, affecting 19 percent of respondents.[14]

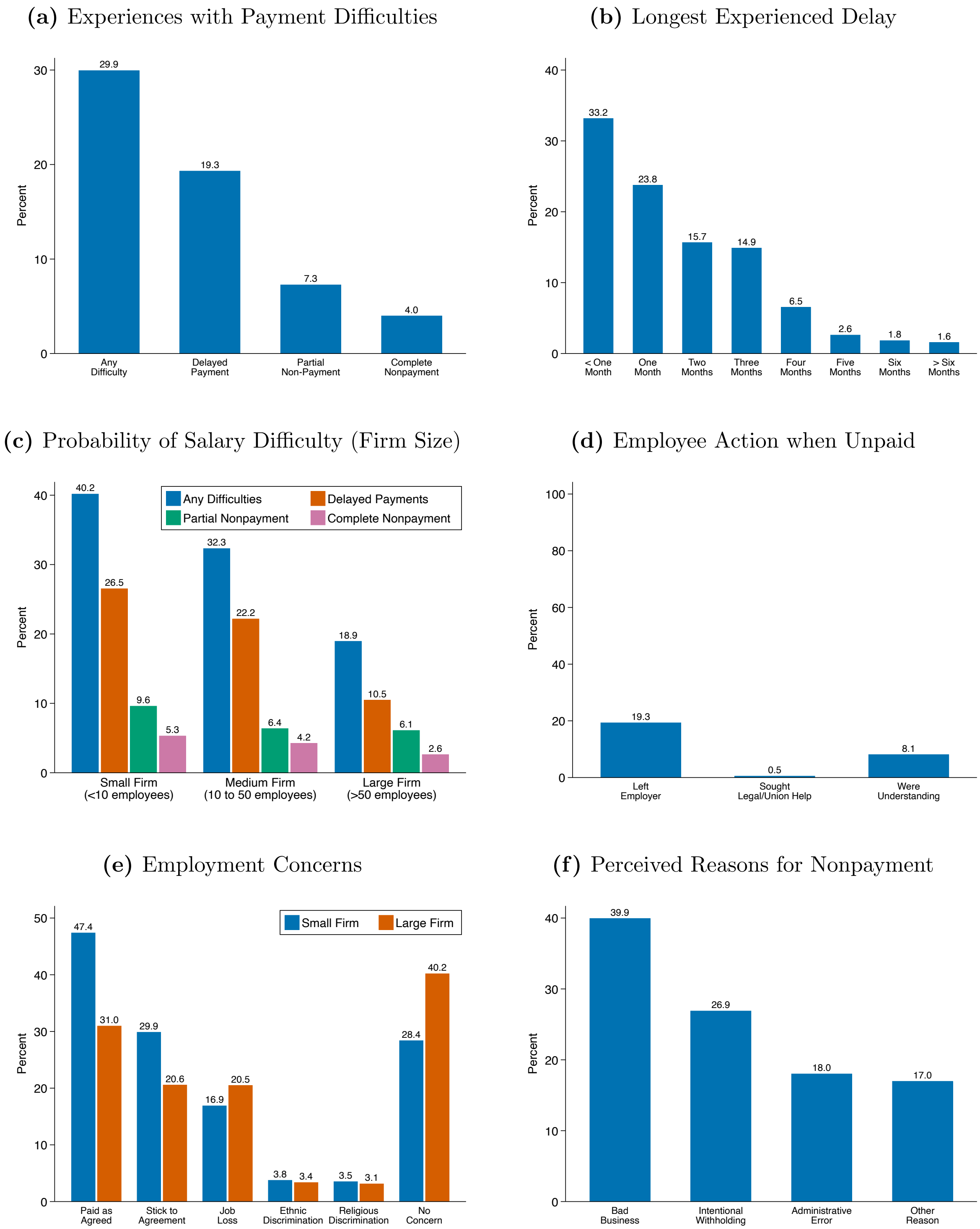
Fact 2: The median delay is one month. Panel (b) shows the distribution of delay durations. Among respondents who experienced payment delays, the median duration of their longest reported delay was one month. About 25 percent of affected workers experienced delays of four months or longer.

Fact 3: Wage withholding occurs across firms of all sizes. Panel (c) shows that delayed and unpaid wages occur across firms of all sizes. We group firms into three size categories. Salary difficulties become less common as firm size increases: 40 percent of respondents reported their worst salary difficulty at small firms (fewer than 10 employees), 32 percent at medium-sized firms ($10 - 50$ employees), and 19 percent at large firms (more than 50 employees). While a large share of cases remains concentrated in small firms, there is still a substantial probability of experiencing salary difficulties even in very large and formal

[13] Respondents could select more than one category, but only very few did.

[14] Appendix Table A.5 shows that individuals expect wage withholding to occur with a very high probability if they were to start working at a firm tomorrow.

Figure 1: Facts About Wage Withholding



Note: This figure presents the six key facts established by our descriptive survey. Panel (a) shows the overall prevalence of wage withholding, and Panel (b) displays the distribution of delay durations. Panel (c) reports the occurrence of wage withholding by firm size, while Panel (d) illustrates employees' reported responses to wage withholding. Panel (e) summarizes workers' concerns about wage employment, and Panel (f) documents their perceptions of why wages were withheld.

firms. This pattern underscores that wage withholding is not limited to the informal sector or small firms but affects the entire economy.[15]

Fact 4: Worker responses are limited. Panel (d) shows how employees respond to wage withholding. Only 19 percent of employees reported leaving their jobs in response to not receiving their salary — a surprisingly low share given that paying agreed wages is a core employer obligation.[16] Firms also appear to face little legal risk, as almost no respondents reported taking legal action or seeking union assistance.[17] Nevertheless, employees disapprove of the practice; only eight percent indicated understanding of the situation.[18]

Fact 5: Wage withholding is a substantial concern. The widespread occurrence of wage withholding translates into strong concerns about being paid as agreed when considering employment. Panel (e) summarizes these concerns, showing that reliability of pay is the primary issue workers associate with both large and small firms.[19] Just over 47 percent of respondents expressed concern about salary payments when thinking about employment at small informal firms, and 31 percent expressed concern about employment at large formal firms.[20] In addition, respondents also expressed concerns about employers not honoring other aspects of the work agreement (e.g., working hours), job loss, and religious or ethnic discrimination in the workplace.[21]

[15] Panel B of Appendix Table A.4 shows that 49 percent of survey respondents know that someone within their direct social environment has experienced salary difficulties.

[16] Panel A of Appendix Table A.3 provides details on how respondents coped financially with unpaid salaries.

[17] Public sector employees also experience wage withholding for example.

[18] In Panel A of Appendix Table A.4, we provide additional evidence that the social norm and clear expectation is for firms to pay their employees fully and on schedule.

[19] Survey participants were asked whether they had any concerns when thinking about employment and, if so, which ones. To avoid biasing responses, they could express more than one concern, so percentages do not sum to 100 but reflect the frequency with which each concern was mentioned.

[20] Because employment in small businesses owned by members of one's social network is common, we also present results from a randomized subset of respondents who were explicitly asked to consider employment at a small business owned by a family member or friend. The level of concern remains largely unchanged.

[21] In Appendix Table A.6 we provide additional descriptive evidence that salary uncertainty influences stated employment preferences. Our survey respondents overwhelmingly express a preference for self-employment over wage employment mirroring synthesized findings by Breza and Kaur (2025).

Fact 6: Workers perceive wage withholding as both involuntary and deliberate. Panel (f) illustrates employees' perceptions of why their wages were withheld. Employees attribute a substantial share of these difficulties either to employers' inability to pay due to poor business conditions (40 percent) or to deliberate wage withholding (27 percent). These responses suggest two main perceived reasons for wage withholding: employers' liquidity constraints and intentional nonpayment.

3 Theoretical Framework

In this section, we develop a theoretical framework to understand the trade-offs workers face when wages remain unpaid and to generate predictions that guide the interpretation of our empirical results.

We study the decision problem of a forward-looking worker who must choose how much costly effort to exert when wages remain unpaid. The worker's effort determines the probability of continued employment, as in Lazear (2000) and Huck et al. (2011). Firms come in two types that differ in payment reliability: a strategic firm never pays wages, while a non-strategic firm pays in full unless hit by a liquidity shock. Upon taking the job, the worker holds an initial belief about the firm's type. Any unpaid balances roll over into the next period and are added to the amount the worker ought to receive. As arrears accumulate, the worker updates beliefs about the firm's type — and thus the expected probability of future payment — and compares the continuation value of staying with the value of the outside option. Unpaid balances therefore create a tension: they reduce the perceived likelihood of eventual payment but increase the potential future payoff. This trade-off is the core insight of the model generating ambiguous effort responses. Our framework, which abstracts from strategic interactions, yields a simple belief structure linking workers' expectations and outside options to effort decisions.[22] We match the model to empirically elicited beliefs about

[22]In Appendix F.5, we extend the framework by endogenizing firms' repayment behavior and show that the simplified version captures the essential worker-side mechanisms without loss of generality: We numerically solve for an equilibrium in which workers' beliefs and firms' repayment policies are jointly consistent.

payment probabilities to derive qualitative predictions for the context we study.

Setup. The model is set in discrete time over an infinite horizon, and firms are of two permanent types: strategic (S) and non-strategic (N). Firms offer wage w to workers, but strategic firms never pay, whereas non-strategic firms pay unless experiencing liquidity shocks. Each period, non-strategic firms face i.i.d. liquidity shocks $L \sim \text{Geom}(\rho)$ that end with probability ρ implying that a shock can persist across multiple periods. During a shock, the firm withholds the wage; when the shock resolves, it resumes full payment.

Workers hold priors over firm types: let λ_0 denote the probability that a newly matched worker believes their employer is of type S. In each pay cycle t, firms either pay the full amount due or nothing — if they are the strategic type or face a liquidity shock. Let $\chi_t \in \{0, 1\}$ denote the realized payment decision applied to both past arrears and the current wage, so that arrears evolve according to $B_{t+1} = (1 - \chi_t)(B_t + w)$.

Workers observe χ_t and update beliefs about the firm's type, and thus the likelihood of future payment, according to Bayes' rule. Beliefs evolve according to

$$\lambda_t(\lambda_0, \rho) \; = \; 1 - \frac{(1-\lambda_0)\, \exp(-\rho B_t/w)}{\lambda_0 + (1-\lambda_0)\, \exp(-\rho B_t/w)}, \quad \lambda_0 \in (0,1),\, \rho > 0, \tag{1}$$

for as long as workers remain unpaid. Here, λ_t denotes the posterior probability that the worker faces a strategic (non-paying) firm after observing no payment ($\chi_t = 0$). After updating beliefs, workers choose costly effort e_t which determines the probability $p(e_t)$ that the match continues, with $p'(e_t) > 0$ and $p''(e_t) \leq 0$. If the match breaks, the firm must repay outstanding wages with probability ϕ.[23]

3.1 The Worker's Problem

The worker's optimization problem can be expressed by the following value function:

[23] All assumptions are formally laid out in Appendix F.1.

$$V(B_t, \lambda_t; \rho) = \max_{e_t} \Big\{ u \overbrace{\big(\chi_t(B_t + w)\big)}^{\text{observed payment}} - \psi(e_t)$$
$$+ \beta \Big[p(e_t) \underbrace{\mathbb{E}[V(B_{t+1}, \lambda_{t+1}; \rho)]}_{\text{expected continuation value}} + (1 - p(e_t)) \underbrace{\big(V^{\text{out}} + \phi u(B_{t+1})\big)}_{\text{expected value if match ends}} \Big] \Big\}. \quad (2)$$

Equation (21) shows that the worker's value depends on two key objects corresponding to the two possible outcomes of the match. When the match breaks with probability $1 - p(e_t)$, the worker receives the outside option and, with probability ϕ, any recovered arrears. We treat ϕ as exogenous, capturing the strength of enforcement: in well-functioning judicial systems, ϕ approaches one, while in environments such as Nigeria, where enforcement is weak, ϕ is likely close to zero.

When the match continues with probability $p(e_t)$, the worker receives the expected continuation value, $\mathbb{E}[V(B_{t+1}, \lambda_{t+1}; \rho)]$. This term captures the central trade-off. As arrears accumulate, the total amount owed increases, mechanically raising the expected utility from remaining with the firm since more pay is due in the next period. At the same time, beliefs about eventual repayment decline with higher arrears, lowering expected utility. The net effect of arrears on the continuation value — and thus on optimal effort — is therefore ambiguous.

3.2 Predictions of the Model

We now use the model to derive three predictions about worker behavior and productivity. Three contextual features of our setting are important for the model. First, the probability of recovering unpaid wages through the legal system is low (see Figure 1d). Second, workers hold a high prior belief that they may face a firm that does not pay wages as agreed (see Appendix Table A.5). Third, even after prolonged nonpayment, workers maintain strong beliefs that they will eventually be paid (see Appendix Figure A.1).

We capture these features through three parameters that reflect workers' beliefs and enforcement conditions in our setting. First, we set the probability of recovering unpaid wages, ϕ, to the share of respondents who reported attempting legal recovery of arrears (Figure 1d). Second, we estimate λ_0 and ρ via nonlinear least squares using workers' beliefs about the likelihood of eventual payment after sustained nonpayment, elicited in the survey described in Section 2.2.[24] Together, these values — $\lambda_0 = 0.41$, $\rho = 0.14$, and $\phi = 0.01$ — capture the key features of the environment described above.[25]

Prediction 1 (Initial Effort Response). Under our parameterization, we expect effort to initially increase when wages remain unpaid. Because the probability of sustained liquidity shocks is high, workers expect that delayed wages will eventually be paid, making continued effort worthwhile in the short run. In contrast, in environments with strong enforcement — such as the United States — where the probability of recovering unpaid wages after separation is high, effort would decline immediately once payments are delayed. This pattern is illustrated in Panel (a) of Figure 2, with the formal derivation provided in Appendix F.2.

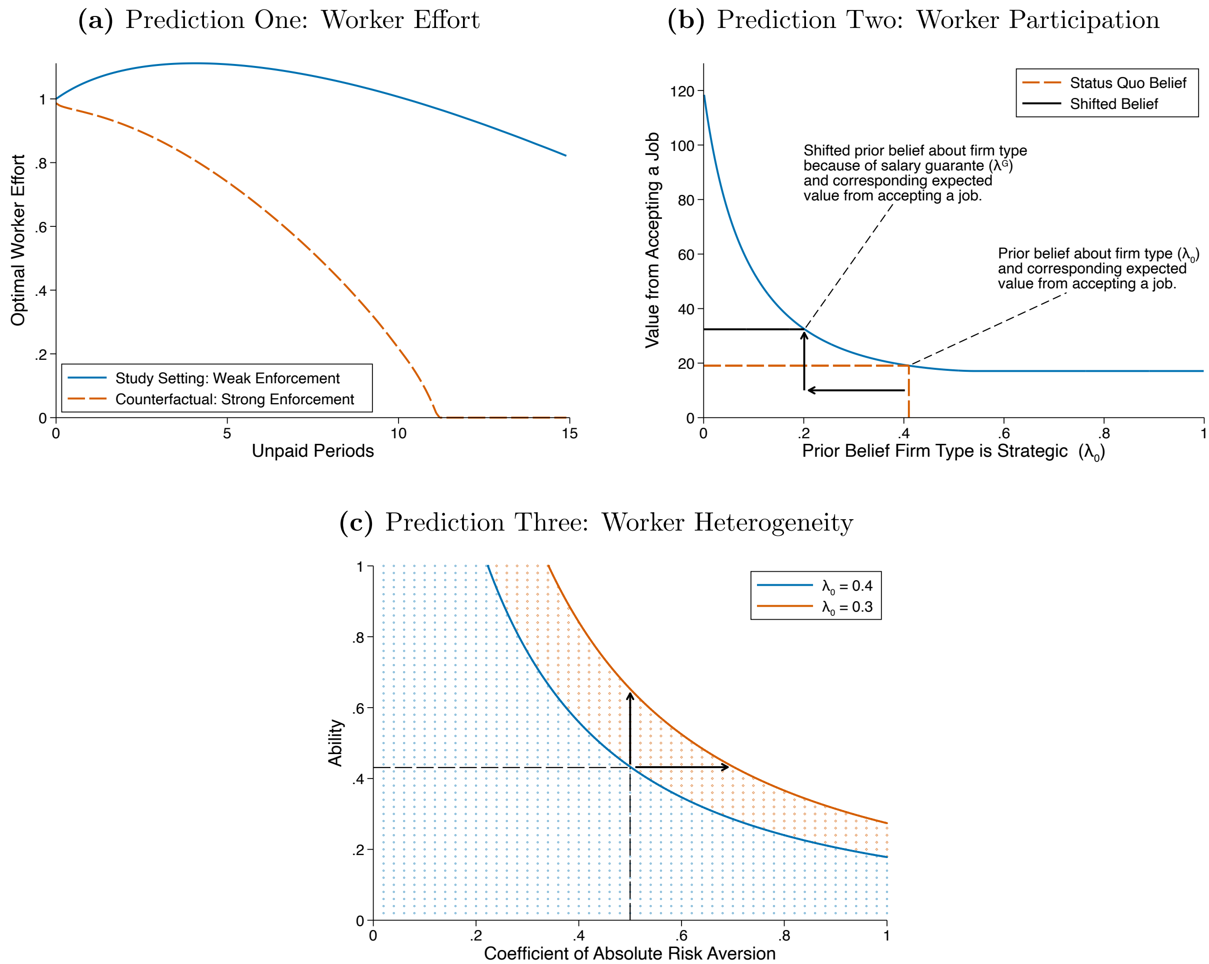
Prediction 2 (Worker Participation). Workers' beliefs influence not only their behavior on the job but also their decision to accept employment. Assume a worker accepts employment if and only if $\mathbb{E}[V(B_0; \lambda_0, \rho)] \geq V^{\text{out}}$, i.e. if its expected value exceeds the outside option. Suppose firms can issue a credible signal G, such as a third-party-verified salary guarantee, that lowers the perceived probability of being matched with a non-paying firm ($\lambda_0^G < \lambda_0$). This guarantee would increase the expected value of the job $\mathbb{E}[V(B_0; \lambda_0^G, \rho)] > \mathbb{E}[V(B_0; \lambda_0, \rho)]$ as Panel (b) of Figure 2 illustrates. Consequently, job take-up should unambiguously rise when offers include a credible signal of firm reliability.

Prediction 3 (Worker Heterogeneity). We now introduce worker heterogeneity along two dimensions: ability and risk aversion. We make two assumptions. First, workers' outside

[24] See Appendix F for details on estimation.

[25] See Appendix F for additional details on model parameterization.

Figure 2: Predictions of the Model



NOTE: This figure plots the three predictions of the model. Panel (a) shows workers' optimal effort when wage payments fail to materialize. The blue solid line illustrates the predicted dynamics under weak contract enforcement, while the dashed orange line represents the case with stronger enforcement. Panel (b) shows how the value from accepting a job varies with the prior belief that the firm will never pay the promised wage (i.e. is of type S). Ex ante, this value declines with higher prior beliefs of facing a non-paying firm; thus, any reduction in the initial prior λ_0 increases the expected value of the job. Panel (c) shows the set of workers — characterized by ability and absolute risk aversion — who would accept the risky job over their outside option. The blue solid line depicts workers accepting the job under the prior $\lambda_0 = 0.41$. Any reduction in the belief of facing a nonpaying firm makes wage employment more attractive. Importantly, marginal entrants can differ along two dimensions: higher ability or greater risk aversion. The orange line illustrates this margin by showing the combinations of worker types that would accept wage employment following a shift to $\lambda_0 = 0.3$.

options increase with ability. Second, the outside option provides a risk-free stream of income.[26] When uncertainty about wage payments is high, accepting wage employment becomes risky relative to this safe but lower outside income. Panel (c) of Figure 2 simulates this trade-off, showing combinations of risk aversion and ability for which wage employment is profitable. The exercise highlights a key insight: lower initial beliefs about firm reliability (λ_0) attract workers who are either more able or more risk-averse, or both.

4 Experimental Design and Implementation

Our experiment is designed to test the predictions outlined in Section 3 by examining three complementary dimensions of salary uncertainty: (i) employees' on-the-job effort responses to delayed salary payments, (ii) the overall impact of salary uncertainty on labor force participation, and (iii) its effect on workforce composition, specifically regarding the types of workers who choose to accept employment.

To examine these outcomes, we incorporated a firm in Nigeria, The Spartak Consult, whose primary business activity is data classification and labeling. As a newly established firm, The Spartak Consult had no preexisting reputation that could influence employee perceptions or behavior. Through this firm, we were able to extend job offers and hire jobseekers for short-term positions while maintaining full control over working conditions and salary payments.

Our experimental design follows a two-stage randomization. First, we designed three distinct job offers explicitly varying in terms of salary certainty; these offers form our initial treatment arms and were randomly assigned to recruitees. To credibly convey differences in salary certainty to jobseekers, we collaborated with a local recruitment agency so that all salary-related information came from a third party rather than directly from the employing firm. Through this collaboration, we recruited individuals through two distinct strategies to reach different populations: individuals proactively responding to job advertisements, and self-

[26] See Appendix F.4 for formal statements.

employed individuals who may not actively seek wage employment. This strategy allows us to assess heterogeneous responses to salary uncertainty across these populations.

In the second stage, conducted in two rounds with varying compositions of the two recruited populations, we hired subsets of interested respondents. Among those who accepted job offers that explicitly mentioned the possibility of delayed payments, we implemented our second randomization by randomly assigning salary delays. To precisely measure productivity responses to these treatments, we developed a job task tailored to this experiment.

The experiment ran from January to October 2025. We posted job advertisements from January to early March and received most responses then, though some continued through early June. In-person recruitment took place in two waves, from January to March and again from June to July. Employment occurred in two rounds: February to May, and August to October. End-of-employment surveys indicate that employees perceived the setup as genuine. In both employment rounds, approximately 80 percent stated that image labeling was the primary purpose of their employment arrangement (see Appendix I.1 for details).

4.1 Job Offer Treatments

Employment Terms Treatment 1 (Control Arm). In the first treatment arm, jobseekers are informed that they work for a local Nigerian firm and receive a fixed monthly salary. No additional information is given to jobseekers about payment modalities of salaries or the firm.

Employment Terms Treatment 2 (Uncertainty Arm). In the second treatment arm, jobseekers received the same information as in treatment arm 1, but were additionally told that each pay cycle carried a probability of salary nonpayment due to Nigeria's difficult economic conditions. This probability was randomized to take values of 10%, 20%, 30%, 40%, or 50%.

Discussing this risk ex ante with jobseekers is a limitation of the design, as wage withholding typically occurs unexpectedly. Disclosure, however, was necessary for ethical reasons. While workers were given a precise probability of nonpayment for each cycle, they received no information about the likelihood or timing of eventual repayment. Thus, if wages were not paid on the scheduled date, employees did not know whether, or when, repayment would occur.

Given the adverse economic context, jobseekers were generally unsurprised by the possibility of nonpayment. As shown in Appendix I.2, receiving this information did not significantly alter individuals' beliefs about timely salary payments.[27]

Employment Terms Treatment 3 (Salary Certainty Arm). In the third treatment arm, jobseekers receive the same information as in Treatment Arm 1. Jobseekers are additionally informed that their salaries are guaranteed to be paid on time. The firm uses a third-party automated payment system, directly connected to a bank account that holds sufficient funds to fully cover salaries for the entire duration of the employment period. There would be no uncertainty regarding timely salary payments. Jobseekers are further reassured that no previous employees receiving this offer have reported delayed or unpaid salaries. This treatment provides the credible salary-certainty shock required to identify labor market participation and workforce composition effects.

4.2 Recruitment and Measure of Take Up

To effectively conduct the recruitment according to our experimental needs, we collaborated with a local recruitment agency, *Unlocking Creativity*. The collaboration allowed some of our enumerators (referred to as recruiters in the following) to be temporarily affiliated with *Unlocking Creativity*, visibly representing the agency during recruitment (e.g., wearing official identification badges and using agency-branded materials). Recruiters then presented themselves as recruiting for their client, *Spartak Consult*, our company set up for this exper-

[27]We interpret this as further evidence that wage withholding is sufficiently common that making the possibility explicit does not meaningfully change expectations about being paid on time.

iment. This approach enabled us to credibly conduct in-person recruitment leveraging the credibility of the agency's credentials while retaining complete control over the recruitment team and their training to ensure maximum compliance to the study protocol. Moreover, this collaboration enhanced the credibility of the guaranteed salary assurances in employment terms treatment arm 3. Recruiters' visible affiliation with an independent agency made their statements about salary payments more credible than if they had come from the client company's own recruiters, enabling us to credibly vary salary certainty.[28]

Job Advertisement Recruiting. To recruit employees actively seeking employment — essential for estimating effort responses to salary delays among typical employees, and serving as a comparison group for workforce composition analysis — members of the field team posted physical job advertisements across selected areas in Lagos from January to early March 2025.[29] The form, content, and placement of these advertisements, shown in Appendix Figure B.2, were chosen to closely resemble local norms.[30] The advertisements were printed in black-and-white on letter (A4) size paper and pasted on walls, poles, and similar public surfaces in selected areas within a roughly one-hour commuting radius of the work locations. They provided only limited information, simply mentioning the general nature of the work (data classification tasks) and a salary range (50,000 NGN to 85,000 NGN), and instructed jobseekers to contact *The Spartak Consult* via phone call or WhatsApp message to express their interest. Providing only basic information in the job advertisements also served our experimental design. By limiting initial details, we obtained an initial sample of jobseekers who responded to the job advertisement independently of any subsequent information about salary payment conditions. All jobseekers who responded to the job advertisement were invited to a job interview.

Interviews were conducted starting on January 27, 2025, either at a rented event hall or at the

[28]Follow-up phone surveys, described in greater detail in Appendix I.2 and Appendix Figure I.3, confirm the effectiveness of this treatment in reducing individuals' concerns about receiving their agreed salary payments.

[29]We started posting job advertisements on January 11, 2025.

[30]Appendix Figure B.3 shows our job advertisement among similar job postings in Lagos.

company's office.[31] During the interviews, we collected jobseekers' baseline characteristics and administered a shortened version of Raven's Progressive Matrices, as well as a choice experiment designed to elicit jobseekers' preferences over different types of job offers.[32]

At the end of each interview, we provided jobseekers with a letter specifying their terms of potential employment. The terms of employment letters differed in terms of the monthly salary offer — randomized between 55,000 NGN and 85,000 NGN (approximately USD 22 to 34) — and also conveyed different information about salary reliability according to the three treatment arms described in Section 4.1.

In-Person Recruiting. To effectively reach our target sample of individuals who are not necessarily looking for employment — and who thus plausibly have higher outside options, crucial for estimating labor force participation and workforce composition effects — we implemented an in-person recruitment strategy. Recruiters frequented pre-specified market areas and public spaces, approaching individuals according to a predetermined skip and selection pattern.[33] Approached individuals were invited to participate in an immediate, approximately 20-minute job information session. For individuals who expressed interest right away, recruiters proceeded with the job information session; those initially hesitant were offered a small monetary incentive (randomized between USD 0.4 and USD 1.5), paid immediately after the session, regardless of their subsequent interest in employment.[34]

The job information session mirrored the previously described job interview; we collected baseline demographic information, administered the choice experiment, and provided the same terms-of-employment letter at the end. As before, the letter stated a monthly salary

[31] Initially, we held interviews on consecutive days to recruit employees for the first employment round which was planned to start in early February. Subsequently, interviews were usually held on a weekly basis. Appendix Figure B.5 reports the days on which we held interviews and the number of interviews per day.

[32] We followed Langener et al. (2022) in designing a shortened version of Raven's Progressive Matrices. Details about the choice experiment are provided in Appendix K.

[33] More details are provided in Appendix B.1.

[34] It was important that the incentive not come directly from the employing company, as this would send a strong signal about the firm. Therefore, the incentive was paid by the recruitment company.

randomized between 55,000 NGN and 85,000 NGN (approximately USD 22 to 34), and varied in salary reliability according to the three treatment arms described in Section 4.1.

In-person recruitment was conducted in two waves. The first wave began on January 20, 2025 and ended on March 6, 2025. The second wave began on June 30, 2025 and ended on July 31, 2025. Appendix Figure B.4 shows the days on which we conducted in-person recruitment and the number of interactions per day.

Measuring Take Up. Our first primary outcome measure is individuals' willingness to accept jobs under the conditions outlined in their respective letters detailing terms of potential employment. To credibly measure actual willingness to work, jobseekers were required to attend an *orientation day* to formally express interest in the offered positions.[35] Attendance at the *orientation day* provides a credible measure of job acceptance, as it constitutes a costly action clearly signaling genuine interest in employment.

4.3 Salary Delay Treatments

To identify employees' on-the-job effort responses to delayed salary payments, we randomly implemented salary delays among employees who accepted job offers in employment terms treatment arm 2. Because these job offers explicitly informed employees about the possibility of salary non-payment in any given pay cycle, we could ethically implement randomized salary delays for employees assigned to this treatment arm. The randomization of salary delays was stratified by age, gender, and treatment probability. Each pay cycle, employees were randomized to either receive their salaries on time or not receive payment, in which case unpaid salaries were added to the balance owed during the next pay cycle. Consequently, an employee could experience salary delays never, once, or multiple consecutive times.[36] Multiple consecutive treatments imply salary delays lasting several pay cycles, with the

[35]The *orientation day* was a pre-scheduled, in-person event designed as a costly signal to elicit jobseekers' interest. Although we provided additional job details during the session, the primary purpose was to record the jobseekers' interest given their terms of potential employment.

[36]We ensured that treatment status could not switch 'on and off' more than once.

maximum consecutive treatment duration restricted to three pay cycles.

While every employee ultimately received their full salary, the wording used in the job offers — "salaries may not be paid in a given pay cycle" — was chosen to maintain high ex-ante uncertainty. This design aimed to closely mimic salary delays at other firms, where the possibility of delays typically remains unannounced until they occur.[37] By comparing the productivity of employees who have outstanding salary balances with those who receive salaries on time, we can estimate the treatment effects of salary delays.

4.4 Job Task

To accurately measure employee productivity, we designed a work task — labeling images of Lego bricks — tailored to the particular requirements of this research project.[38] This task represents a specific instance of data labeling, an increasingly common type of work in which workers manually add informational labels to individual data points (in this case, images) to identify their attributes or assign them to specific categories. Such labeled data serve as ground-truth datasets essential for training artificial intelligence models — for instance, image-recognition tools.[39]

We digitally created 44,000 images of Lego bricks and defined six distinct categories — related to the color, type, and quantity of bricks visible in each image — according to which employees had to label the images. Categories could require one or multiple labels, resulting in a total of 10 label observations per employee for each image.[40] Employees performed these labeling tasks individually on provided work computers using the online data labeling

[37] While salary delays typically occur unexpectedly at other firms, we explicitly disclosed this possibility in our job offers for ethical reasons. See Section 4.1 for a discussion of why we believe that announcing these possibilities in the job offers does not distort our estimates.

[38] We acknowledge that labeling Lego bricks is not a typical job task in this context. However, this task closely resembles skills required in many jobs: carefully following instructions, managing tasks of varying difficulty, and maintaining productivity in repetitive assignments.

[39] We provide context about this type of work and show that it is a common task in Appendix C.4.

[40] For example, one category required employees to label the (uniform) color of the Lego bricks in the image, selecting a single label from a predefined set of colors. Another category asked employees to indicate which specific types of Lego bricks appeared in the image, requiring separate "yes" or "no" labels for each of five possible brick types. More details on the labeling categories are provided in Appendix C.2.

platform *Labelbox*, which is designed for exactly this type of work. After completing all labels for an image, the next image appeared immediately (Appendix Figure C.3 shows an example). Importantly, all employees labeled the same dataset of images in identical order, ensuring that every image was labeled by each worker. This design eliminates variation in task content as a source of performance differences and allows us to account for learning over time when estimating treatment effects.

We designed the labeling task to include two clearly distinguishable types of images: 'regular' and 'high-stakes'. Approximately 91 percent were regular images with plain white backgrounds. The remaining nine percent, designated as high-stakes, featured visually noisy, captcha-style backgrounds (an example is shown in Appendix Figure C.2), increasing their labeling difficulty. During their initial training, employees were explicitly instructed to devote additional care and effort to labeling these high-stakes images, as mistakes would be particularly costly for the company.

This pre-registered design allows us to assess effort responses on a generic task (regular images) and on a more demanding, high-stakes task where employees also have the opportunity to retaliate against the company. If outstanding salary balances induce retaliation, we would expect performance on high-stakes images to deteriorate relative to regular images, as mistakes on these images are particularly costly for the employer. By contrast, if employees intend to demonstrate effort to their employer, performance on high-stakes images should be at least as high as on regular images, given the extra effort. Further details on the image generation process and labeling categories are provided in Appendix C.

Measuring Productivity. This job task allows us to evaluate employee performance at the individual-image level, our key outcome measure. Because we digitally generated all images, we know the correct labels for each image's categories and can directly compare employee-assigned labels to the correct ones. Our main analysis relies on two complementary, pre-registered productivity measures. The first is a continuous index measuring the total number of correct labels per image, ranging from zero (no correct labels) to ten (all labels correct). The second is a binary index equal to one if all labels on an image are correct and zero otherwise. For robustness, we also consider a productivity measure based on a pre-registered subset of the labeled images (see Appendix E.1).

4.5 Employment Arrangements

Employment took place across two distinct rounds, each designed to identify different effects. The first employment round primarily focused on estimating employees' on-the-job effort responses to salary delays. For logistical reasons, this round began in two batches, starting on February 3 and February 10, 2025.[41] The second employment round, starting on August 4, 2025, was designed to assess productivity differences between worker types and further evaluate effort responses among employees with higher outside options.

Employees received instructions to resume at one of six work locations. All work locations were in proximity to each other accommodating 24 to 100 individuals. Employees were expected to work 7.5 hours per day with a 30-minute break scheduled around lunchtime. During work hours, employees only task was to label the images of Lego bricks as described in Section 4.4

On their first workday, employees received extensive training, covering computer use and detailed instruction on the labeling task, as well as logistical aspects of the job. The exact dates for the scheduled biweekly disbursements of the monthly salary were communicated to employees. Additionally, each work location had a fixed seating arrangement, and employees were allocated to tables of four.[42] Communication between tables was explicitly prohibited. To enforce this rule as much as possible each table also had its own schedule, specifying start times, breaks, and closing times. These measures were implemented to address potential

[41]Individual employees effectively started on different dates if they missed their assigned first workday or if they replaced another employee who quit shortly after starting.

[42]In a few exceptions the number of employees per table was three or five.

spillover effects of salary delays and are discussed and validated in Appendix E.2.

4.6 Supplementary Data Collections and Interventions

In addition to our primary data, we collected supplementary datasets from employee and recruitee samples, in some cases through additional interventions. These data provide further insights into employees' experiences and responses to wage withholding, recruitees' perceptions of job offers across the three treatment arms, and help contextualize the magnitude of the treatment effects. Specifically, we: (i) administered an end-of-job survey to elicit employees' beliefs about salary delays and job perceptions; (ii) implemented a bonus-payment intervention in the final week of employment round one to benchmark the magnitude of effort responses; and (iii) ran a follow-up survey with recruitees who were not hired to capture their perceptions of employment terms across treatment arms. Full details and results are reported in Appendix I.

4.7 Ethical Review

This research underwent ethical review in both the United States and Nigeria and received IRB approval from the relevant institutions. We recognize that experimentally delaying salary payments raises important ethical considerations; however, we consider the study ethically appropriate for several reasons. First, we communicated the possibility of delayed wages clearly and transparently to all affected participants. Job offers explicitly stated that salaries might not be paid in a given pay cycle, and participants voluntarily chose whether to work under these conditions. Second, the phenomenon we study — wage withholding — is common in the study setting, as documented in this paper, meaning participants were not exposed to risks beyond those encountered in everyday life. Third, we conducted a full debriefing with all participants, who received complete information about the research and were paid any outstanding salary balance, including interest to account for the delay period. Finally, the study contributed directly to participants' welfare: we paid over USD 100,000 in wages in this setting, creating temporary employment opportunities for 600 workers.

5 Data and Descriptive Statistics

5.1 Characteristics of Recruitment Samples

From January 2025 to July 2025, we recruited three distinct samples of individuals to estimate both workers' responses to delayed wages and the impact of salary uncertainty on labor-force participation. Because we recruited for real jobs, these recruitees also formed the pool of potential employees, and we later hired a subset of respondents from these samples for our two rounds of employment. First, we interviewed 638 respondents who directly applied to our posted job advertisements (Sample I).[43] Separately, recruiters interviewed a total of 1,079 individuals through our targeted in-person recruitment strategy.[44] Among these, 258 individuals agreed to participate immediately without incentive (Sample II), while 821 only agreed after receiving the monetary incentive (Sample III). Table 1 summarizes sample characteristics.

As anticipated, the three samples differ noticeably along certain dimensions. While demographic characteristics are reasonably similar, self-reported job-search status differs substantially across groups. The share of self-employed individuals is also substantially higher among those who required a monetary incentive to participate in the job information session.[45] Attendance at the orientation day also differs noticeably across samples. While attendance was 95 percent for individuals recruited through job advertisements (Sample I), it was lower for those recruited in-person — 60 percent for the unincentivized group (Sample II) and 44 percent for the incentivized group (Sample III). The highest level of schooling differs substantially across samples. Among job-advertisement recruitees, 39 percent hold

[43] Appendix Figure B.5 shows the timing of the job-advertisement recruiting.

[44] Appendix Figure B.4 shows the timing of the in-person recruitment; recruiters' conversion rate from approach to interview was 44 percent.

[45] We refer to this group broadly as self-employed, although it includes individuals running small-scale businesses—often not categorized as self-employment by respondents—and those engaged in casual daily labor, which respondents frequently did not perceive as employment at all. For further discussion, see Barker et al. (2024).

Table 1: Characteristics of Recruitment Samples

	Sample I	Sample II	Sample III
Characteristics			
Age	29.53	25.65	27.16
% Female	0.53	0.40	0.42
Raven's Matrices	8.16	8.09	7.40
% Searching for a Job	0.99	0.89	0.62
% Attending Orientation	0.95	0.60	0.44
Employment Status			
Unemployed	0.51	0.47	0.26
Employed	0.10	0.11	0.15
Self-Employed	0.39	0.41	0.59
Highest Schooling			
Secondary School	0.39	0.60	0.63
Vocational Training	0.01	0.02	0.01
National Diploma	0.16	0.15	0.13
University	0.44	0.22	0.20
Observations	638	258	821

NOTE: This table reports average characteristics across our three samples of recruitees. Sample I: job-advertisement recruitees, i.e. individuals who responded to our job advertisement. Sample II: in-person recruitees who participated in the job information session without a monetary incentive. Sample III: in-person recruitees who required a monetary incentive to participate in the job information session. Appendix Table B.8 provides a more detailed version of this table, reporting additional variables, standard deviations, and percentiles.

only a high-school degree, compared with 60 and 63 percent in the two in-person recruited samples. Consequently, the pattern reverses for postsecondary education: 44 percent of job-advertisement recruitees obtained some postsecondary education, compared with 20 and 22 percent in the in-person samples.

5.2 Employment Patterns and Productivity

This section summarizes the composition of the two employment rounds and describes employee performance.[46] The first employment round focused on identifying employees' effort responses to salary delays. We hired 300 jobseekers from Recruitee Sample I (job-ad sample),

[46] Appendix Table B.9 reports employee demographics.

assigning 33 employees to the control group (treatment arm 1) and 267 (about 90 percent, as specified in the pre-analysis plan) to the salary-uncertainty group (treatment arm 2), as shown in the first column of Panel A of Table 2.

Table 2: Employee Samples

	Employment Round 1	Employment Round 2			
	Overall	Overall	Sample I	Sample II	Sample III
Panel A: Employment Sample Composition					
Arm 1: Control	0.11	0.14	0.37	0.00	0.00
Arm 2: Certainty	0.00	0.61	0.63	0.62	0.58
Arm 3: Uncertainty	0.89	0.26	0.00	0.38	0.42
Panel B: Employee Work Patterns					
Images labeled per day	255.73	207.82	212.31	204.31	205.70
Absent (0,1)	0.21	0.18	0.17	0.20	0.17
Hours at work	7.16	7.13	7.16	7.08	7.12
Hours spent working	6.12	6.15	6.17	6.10	6.16
Images labeled in total	15,332.38	8,305.79	8,485.84	8,164.52	8,220.82
Correct labels per image (0,10)	8.37	8.38	8.38	8.43	8.36
All labels correct (0,1)	0.10	0.10	0.10	0.10	0.10
Seconds spent per image	89.96	107.64	103.89	109.81	109.81
Observations	300	300	110	65	125
Employment duration (months)	3	2	2	2	2

NOTE: This table reports employees' performance and terms of employment across the two employment rounds. For the second round, all statistics are presented separately by recruitment sample: job-advertisement recruitees (Sample I), in-person and unincentivized (Sample II), and in-person and incentivized (Sample III). All employees hired in the first round were recruited through job advertisements. Panel A reports the composition of employment terms under which employees were hired. Panel B presents performance-related statistics for employees.

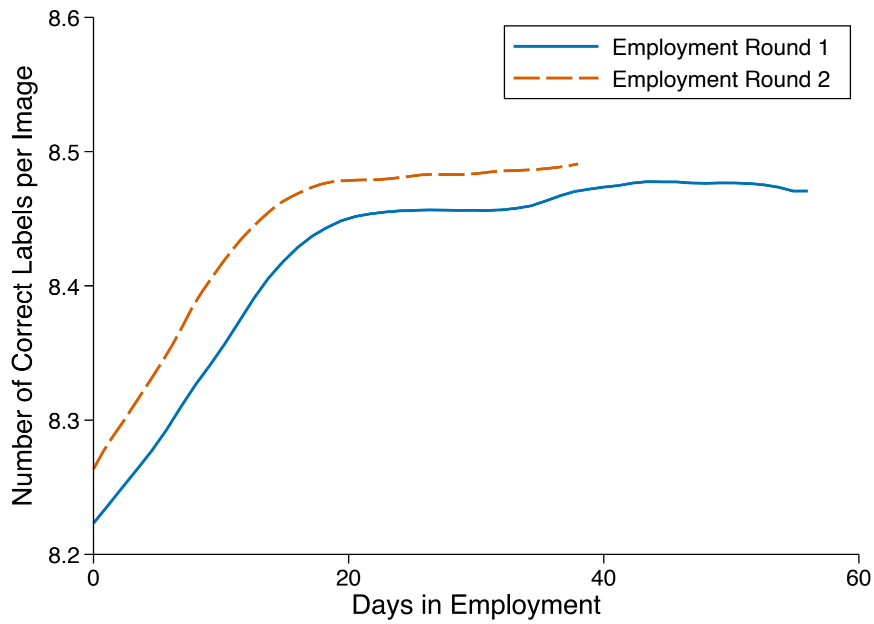
The second round primarily examined workforce composition effects and allowed us to estimate complementary effort responses among employees with higher outside options. We again hired 300 jobseekers—110 from Sample I, 65 from Sample II, and 125 from Sample III. Columns 2-5 of Panel A in Table 2 show that 14 percent (41 employees) received control offers (arm 1), 61 percent (182 employees) received salary-uncertainty offers (arm 2), and 26 percent (77 employees) received salary-certainty offers (arm 3).[47]

Panel B of Table 2 presents descriptive work patterns, which are similar across the two

[47] Percentages sum to 101 due to rounding.

employment rounds.[48] Employees labeled an average of 256 images per day in the first round and 208 in the second. Some employees stopped attending work, and we classify all missed days as absenteeism. Including these cases, absenteeism averaged 21 percent in the first and 18 percent in the second employment round. Conditional on attendance, employees worked an average of 7.16 hours per day in the first round and 7.13 in the second, of which 6.12 and 6.15 hours, respectively, were spent actively labeling images. Employees differed substantially in both speed and accuracy. On average, workers labeled 15,332 images over the course of the first round and 8,305 in the second (reflecting the shorter employment duration). The average employee answered 8.37 out of 10 labeling questions correctly in the first round and 8.38 in the second, though many made minor errors per image, resulting in an overall rate of flawlessly labeled images of about 10 percent.

Figure 3: Learning Curve



Note: This figure plots employees' empirical learning curves by employment round. The outcome is the average number of correct labels per day, smoothed using a local polynomial regression with an Epanechnikov kernel and a bandwidth of three days. The first employment round lasted roughly three months (around 60 working days), and the second lasted about two months (around 40 working days).

[48]We provide summary statistics at the image-level in Appendix C.5.

Figure 3 shows employees' empirical learning curve. After a steep initial increase at the start of the employment spell, performance remained relatively constant over the remainder of the period. The learning pattern looks remarkably similar across the two employment rounds.

6 Empirical Strategy and Results

We now turn to presenting our empirical analysis and experimental findings. We begin by analyzing employees' responses at the intensive margin — that is, examining how salary delays affect the on-the-job effort and productivity of employees who continued working despite experiencing withheld wages. Subsequently, we present the extensive margin results, addressing labor force participation decisions and workforce composition effects. The empirical strategy follows our pre-analysis plan, which pre-specified the outcomes, estimation approach, and aggregation level of treatment effects.

6.1 Intensive Margin: On-the-Job Effort Responses

Effort Response Estimation. To estimate the effect of delayed payments on employee productivity, we exploit random treatment timing, analyzed in Roth and Sant'Anna (2023) and implement the event study estimator of Borusyak et al. (2024). This estimator allows for heterogeneous treatment effects, which conventional TWFE eventstudy estimators do not (Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2022; Roth et al., 2023). The estimator first fits a TWFE regression using only untreated observations and then uses the estimated parameters to predict counterfactual outcomes for treated observations. It is particularly well suited to our setting because the design of our work task allows us to condition estimation on the labelled image q, yielding granular predictions for each image.

Formally, let $G_i \in \mathcal{G} = \{1, 2, \ldots \infty\}$ denote employee i's first treatment date t ($G_i = \infty$ if employee i is never treated). Additional, let $y_{itq}(g)$ denote the potential productivity on image q for individual i on date t when treatment started at g. We can then define event time as $l_{it} = t - g_i$.[49] Event time is measured in workdays since treatment. Define the set of untreated and treated employee-day observations as $\mathcal{S}_0 = \{(i,t) : l_{it} < 0\}$ and $\mathcal{S}_1 = \{(i,t) : l_{it} \geq 0\}$, respectively. We estimate

$$y_{itq} = \alpha + \delta_q + \kappa_i + \mu_t + \varepsilon_{itq} \tag{3}$$

via OLS on $(i,t) \in \mathcal{S}_0$ (i.e., untreated observations only). Here, y_{itq} is the outcome of interest — the continuous or binary index, described in Section 4.4 — for image q, labeled by individual i on day t. The parameters δ_q, κ_i, and μ_t denote vectors of image, individual, and date fixed effects, respectively. For treated observations $(i,t) \in \mathcal{S}_1$, we obtain predicted untreated outcomes $\hat{y}_{itq}(\infty) = \hat{\alpha} + \hat{\delta}_q + \hat{\kappa}_i + \hat{\mu}_t$ and compute the treatment effect of salary delay at the employee-image level as $\tau_{itq} = y_{itq} - \hat{y}_{itq}$. We then aggregate these effects for each event time l using flexible weights $\omega_{itq}^{(l)}$.[50] Formally, we define, with $\mathcal{S}_l \equiv \{(i,t,q) : l_{it} = l\}$,

$$\tau_{l,\omega} = \sum_{(i,t,q) \in \mathcal{S}_l} \omega_{itq}^{(l)} \, \tau_{itq}^{(l)}. \tag{4}$$

We consider two weighting schemes in our analysis. First, each image-level observation receives the same weight, so the estimand is the average treatment effect for event-time g across all images; this is the weighting we apply throughout our main analysis. However, this weighting scheme does not account for differences in employees' labeling speed, as employees who label more images contribute proportionally more to the estimated effect. To address this, we implement a second weighting scheme that assigns equal weight to each employee-workday combination corresponding to event time g. Each image is weighted inversely proportional to the number of images labeled by that employee on that workday. We then

[49] We can then define a binary treatment indicator $D_{it} \in \{0,1\}$ as $D_{it} \equiv \mathbb{1}\{l \geq 0\}$.

[50] We estimate treatment effects at the daily level, as specified in our pre-analysis plan, because we considered this frequency best suited to capturing heterogeneity in work performance before further aggregating treatment effects. For example, performance may be influenced by daily factors such as extreme heat, power outages, or network interruptions.

aggregate the daily treatment effects to obtain weekly treatment effects and an overall ATE.

We compute standard errors in two ways. First, we use the conservative variance estimator proposed by Borusyak et al. (2024) to obtain standard errors for our daily event-time estimates $\tau_{l,\omega}$. We then obtain standard errors for the weekly and overall ATE aggregates by applying the delta method to our daily event-time estimates. Second, we calculate and report bootstrap standard errors, as proposed by Liu et al. (2024), for our final estimates. Since we analyze multiple productivity measures we compute sharpened q-values following Benjamini et al. (2006) to account for multiple hypothesis testing.[51] We report both conventional p-values and sharpened q-values in the results.

Identification of the treatment effects relies on random treatment timing, stratified by age, gender and job offer details. Formally, let J_i denote the stratum of individual i.[52] Let $\{Y_{itq}(g)\}_{t,g\in\mathcal{G}}$ collect all potential outcomes of individual i under all possible treatment start cohorts.

Assumption E1 (Random treatment timing within strata.)**.** *Treatment timing is randomly assigned within strata so that*

$$G_i \perp\!\!\!\perp \{Y_{it}(g)\}_{t,g\in\mathcal{G}} \,\big|\, J_i, \quad \text{and} \quad \Pr(G_i = g \mid J_i = j) = \pi_{jg} \;\; \forall \;\; j \in \mathcal{J},\, g \in \mathcal{G},$$

with $\pi_{jg} \geq 0$.

In words, this assumption states that conditional on the stratification variables J_i, the treatment timing g_i is independent of all potential outcomes, and all individuals in the same stratum share the same assignment probabilities $\{\pi_{jg}\}_{g\in\mathcal{G}}$. This directly implies parallel

[51] This method is suitable for positively dependent p-values, see Anderson (2008) for a discussion. This assumption is reasonable in our setting, as different productivity measures for the same image are likely correlated through underlying worker ability and effort on that particular image.

[52] Strata are defined by gender, age blocks (18-24, 25-34, 35-44 and 45 or older), and job offer details (treatment arm 1 to 3, and for treatment arm 2 the specified probability that salary-nonpayment may happen in a given pay cycle).

trends.[53]

As a validity check, we test for pre-trends following Borusyak et al. (2024), testing whether all pre-treatment coefficients are equal to zero. Full details of the test are provided in Appendix D.

To ease interpretation of the estimated effects, we impose two sample restrictions in all estimations. First, we do not allow treatment status to switch "on and off." To ensure that truthful information was conveyed to employees, individuals were randomized into treatment in each period. This implies that, in principle, an employee could be treated in one period, untreated in the next, and treated again thereafter. To prevent such switching, we restrict assignment so that individuals can only receive treatment in consecutive periods. Once an employee is randomized out of treatment after having been treated in one or more consecutive periods, they are no longer eligible to be randomized into treatment again. In addition, we exclude employees who were treated in one period but then became untreated in the following period. Since initial treatment was randomized, this exclusion effectively removes a random subsample from some weeks of the analysis. The restriction ensures, however, that we only use completely untreated individuals (observations with $l_{it} < 0$) to estimate the fixed effects employed for imputation. Second, we exclude the first 500 labeled images from all estimations. As shown in Figure 3, employee performance exhibits a steep initial learning curve, making early labeling decisions particularly noisy. Excluding these observations does not substantively alter the results and, if anything, leads to slightly more conservative estimates.

Estimating Absenteeism and Day-Specific Outcomes. In addition to productivity measured at the image level, we also estimate absenteeism and other outcomes defined at the workday level, such as total time spent labeling and total work time. Our estimation strategy follows the approach outlined above, again using the imputation estimator of Borusyak et

[53] See Appendix D for a formal argument.

al. (2024), with the difference that outcomes are defined at the day rather than the image level.[54]

Effort Results. We find that salary delay increases effort in image classification by about 0.5 percent relative to the control mean, with the dynamic patterns closely resembling the simulations from our theoretical framework. Figure 4 presents weekly event-study estimates for both the continuous (Panel (a)) and binary indices (Panel (b)), obtained by aggregating the daily treatment effects from equation (4) to the weekly level.[55] Over the course of the experiment, employees increased their effort in response to longer salary delays. The pre-treatment trends in both panels are flat and close to zero. For the continuous index, there is an increase in the final week before treatment, but the coefficient is statistically insignificant at conventional confidence levels.

Table 3 aggregates the daily treatment effects from equation (4) into a single ATE. Columns (1) and (2) present results for the continuous and binary indices, respectively, for the full sample of images. Columns (3) and (4) are analogous but restrict the sample to the 'high-stakes' images, for which employees were encouraged to exert extra effort.[56] The treatment effects are very similar across the two image samples, indicating that employees who experienced salary delay maintained a consistently higher level of effort than those who received their salary on time.

The coefficient on the continuous index quantifies the change in correct labels per image, while the coefficient on the binary index quantifies the change in the probability of flawless labeling. For example, the coefficient in column (1) indicates that employees' performance increases by 0.045 correct labels per image, on average, during periods of payment delay.

[54]Specifically, we estimate

$$y_{it} = \alpha + \kappa_i + \mu_t + \varepsilon_{it} \tag{5}$$

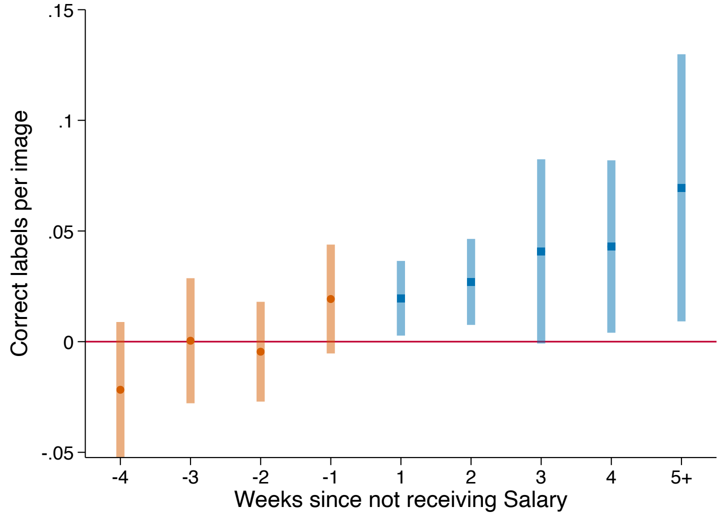
via OLS on $(i,t) \in \mathcal{S}_0$. Here, κ_i and μ_t are defined as before, and y_{it} denotes workday-specific outcomes.

[55]As pre-registered, we considered this the most appropriate level ex-ante for analyzing dynamic productivity responses to salary delay.

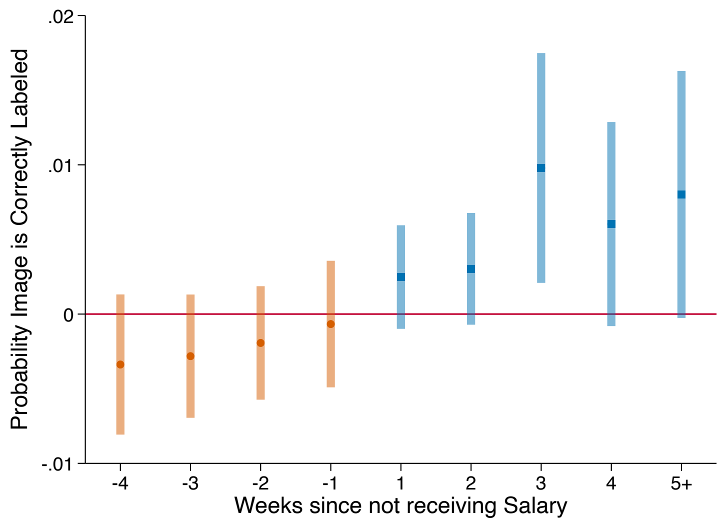
[56]See Section 4.4 for a more detailed discussion.

Figure 4: Event Study Graphs: Estimates from Equation (4)

(a) Effect on Number of Correct Labels



(b) Effect on Probability of Flawless Completion



NOTE: This figure plots the dynamic effect of being owed salary payments on employees' effort measures. Panel (a) presents results for the number of correct labels per image, while Panel (b) shows results for the binary indicator of whether an image was labeled entirely correctly. The blue squares show the estimates of equation (4) aggregated at the weekly level, standard errors are calculated using the conservative variance estimator proposed in Borusyak et al. (2024). The orange circles show estimates from a separate regression testing for pre-trends as suggested by Borusyak et al. (2024) (these are not from the same estimation and coefficients are not relative to a single omitted time period as is often the case in figures like this one), see Appendix D for details. Standard errors are initially clustered at the individual level and aggregated using the delta method. The bars show 95 percent confidence intervals.

Relative to a baseline mean of 8.42 correct labels, this corresponds to a productivity increase of 0.53 percent. This, in turn, corresponds to an increase in the probability of flawless labeling of 0.6 percentage points, as shown in column (2). Relative to an 11 percent baseline probability of flawless labeling, this implies a treatment effect of 5.4 percent.

Table 3: Treatment Effects of Wage Withholding on Worker Effort

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.045	0.006	0.048	0.008
	(0.016)***	(0.003)**	(0.016)***	(0.003)***
	[0.018]***	[0.003]**	[0.019]***	[0.003]***
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Individual	Individual	Individual	Individual
Mean of Dep. Var.	8.43	.11	8.41	.11
Q-Value	0.010	0.016	0.011	0.011

NOTE: This table reports the effect of delayed salary payments on employees' effort measures estimated using equation (4) and aggregated into a single ATE. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the individual level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses.

To contextualize these magnitudes, we compare them to a one-standard-deviation change in performance for each index. Across employees, the standard deviation of performance is 0.39 for the continuous index and 0.04 for the binary index (Panel C of Table 2). These estimates imply that, in the initial weeks of experiencing salary delay, employees improve performance by approximately 11.5 to 15.5 percent of a standard deviation, depending on the productivity measure considered.[57]

As a second benchmark, we compare the estimated performance increase to the increase induced by a bonus payment. In a separate intervention, we offered a bonus of ten NGN per correctly labeled image, which increased performance by 1.4 percentage points. Under the assumption of a linear relationship between bonus amount and performance, the estimated treatment effect of salary delay is equivalent to offering a bonus of approximately five NGN per correct image.[58] Appendix I.3 provides details on the implementation of the bonus pay.

The Appendix contains multiple versions of Figure 4 and Table 3 incorporating a range of robustness checks and complementary results. First, Appendix E.2 addresses potential spillover effects of salary delay across employees — arising from the concern that individuals with different treatment status may share the same work location — and shows that the results are not sensitive to accounting for such spillovers. Second, we show that the results are robust to using an alternative productivity index based only on a pre-registered subset of the labels in Appendix E.1. Third, Appendix E.3 reports results using the alternative weighting scheme for calculating daily treatment effects $\tau_{g,\omega}$, outlined earlier in this section, and shows that results are largely robust to the choice of weighting. Fourth, we find that the results are not meaningfully affected by the level of treatment penetration within a work location, as reported in Appendix E.4. Fifth, we show that treatment effects are larger in magnitude when the sample is restricted to the most difficult images — those with the lowest average labeling performance in the untreated sample. Given high overall labeling performance, additional effort should matter more for the most difficult images, and Appendix E.5 provides evidence consistent with this expectation.

Heterogeneity of Effort Responses. To understand what drives the effort responses to salary delay and if firms may face retaliation from a subset of employees, we investigate

[57]The experiment cannot speak to any results beyond the experimental six-week window.

[58]Five NGN corresponds to around 1.2 percent of the hourly minimum wage.

heterogeneity of treatment effects. Specifically, we estimate versions of Equation (4) in which estimands are separately aggregated for binary subgroups. The results are presented in Figure 5. We examine heterogeneity by subgroups that differ in outside options — and thus in the cost of forgoing unpaid wages or losing their jobs — focusing on pre-treatment performance, postsecondary education, and gender.

Employees' pre-treatment performance is informative about their outside options. Lower-performing workers are less productive and therefore less attractive to alternative employers, limiting their outside options. Similarly, we interpret postsecondary education as a proxy for higher outside options, as better-educated workers are more likely to find alternative employment. In addition, men may have better outside opportunities than women in this labor market, since much of casual employment involves physically demanding work that limits women's access to these jobs. Following the logic of our theoretical framework, groups with lower outside options should increase their effort more strongly in response to withheld wages. These workers are more dependent on their current job, and forgoing outstanding salary balances is more costly for them. They therefore have stronger incentives to intensify effort when salary payments are delayed. At the same time, workers with higher outside options could be considered more likely to retaliate over unpaid wages.

We find evidence consistent with the prediction that effort increases should be particularly pronounced among workers with low outside options while not finding any evidence of retaliation among workers with higher outside options. First, workers who performed below the median before treatment increased their effort significantly more while wages were outstanding than those who performed above the median ($p < 0.01$). Panel (a) shows the dynamic treatment effects: labeling performance increases significantly for below-median performers, while we find a precise null effect for those above the median. Second, splitting the sample by postsecondary education yields similar results. Workers without postsecondary education show a significantly stronger effort response than those with postsecondary education

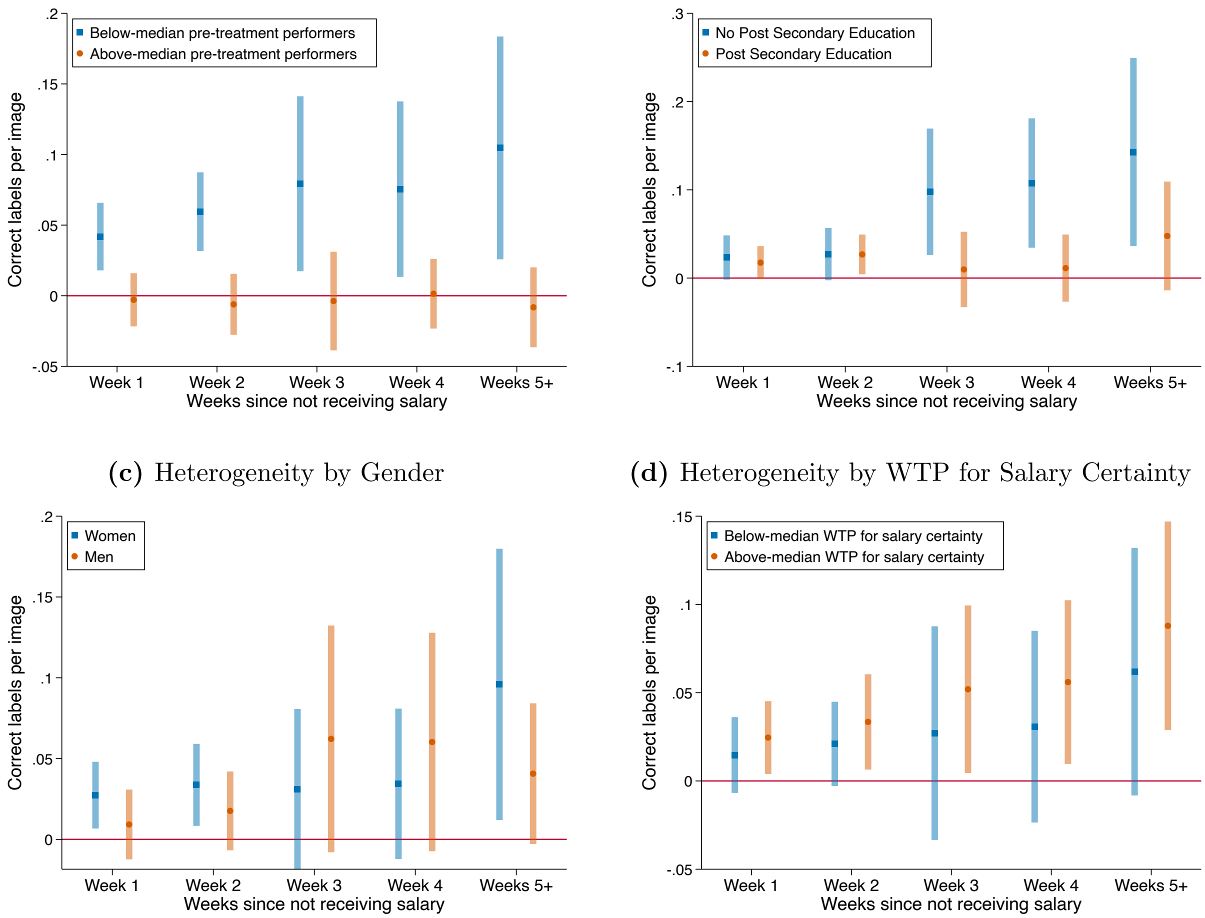
Figure 5: Treatment Effect Heterogeneity

(a) Heterogeneity by Pre-Treatment Performance

(b) Heterogeneity by Schooling

(c) Heterogeneity by Gender

(d) Heterogeneity by WTP for Salary Certainty



NOTE: This figure illustrates treatment effect heterogeneity. The points plot weekly averages of the estimated eventstudy coefficients for the continuous accuracy index, and the bars show 95% confidence intervals. Panel (a) presents heterogeneity by pre-treatment performance, splitting the sample into above- and below-median performers. Panel (b) shows heterogeneity by post-secondary schooling status. Panel (c) reports heterogeneity by gender, and Panel (d) illustrates heterogeneity by WTP for higher salary certainty.

($p < 0.05$). Dynamic results are presented in Panel (b). Third, we find suggestive evidence that women increase their effort slightly more than men in response to wage withholding, but this difference is not significant. Panel (c) displays the dynamic results.

Next, we examine whether heterogeneity in workers' valuation of salary certainty matters for their responses to unpaid wages. To do so, we draw on the discrete choice experiment conducted during the job interviews and information sessions. In the choice experiment, job-seekers made incentive-compatible choices between hypothetical jobs, allowing us to estimate a mixed logit model, and derive individual-level WTP values for higher salary certainty.[59] We may think that different valuations for higher salary certainty transaltes into heterogeneous responses when wages are delayed. However, we find little evidence in favor of this hypothesis. Panel (d) plots the dynamic responses for below- and above-median valuation subgroups. While workers with higher WTP for salary certainty exhibit a slightly more pronounced effort response, we cannot reject equality of responses on average. Importantly, neither group retaliates against the employer.[60]

Absenteeism and Hours Worked Results. Table 4 shows estimation results for absenteeism and related outcomes aggregated by pay cycle. All standard errors are clustered by individual employee and combined using the delta method. Column (1) shows the estimation results for the effect of salary delay on being absent for an entire workday. The coefficient of 0.026 implies that employees were 2.6 percentage points more likely to be absent during the time (two weeks) of the pay cycle in which they did not receive their salary for the first time. This effect, however, is not significant at a reasonable confidence level.

Column (2) shows the effect for total work hours, conditional on showing up to work. Employees have fixed start and end time each day, which correspond to a workday of 7.5 hours. Employees could, however, resume late or leave early reducing the effective number of work hours. Some employees do so, as the average number of work hours is 7.19. We can infer this time from the first and last image that employees label. Column (2) shows that there is no significant effect on total work hours caused by salary delay, conditional on showing up to work.

[59]Details on the methodology and setup of the choice experiment, and the distribution of individual level WTP values are provided in Appendix K.

[60]For all considered outcomes, Appendix Table E.11 reports subgroup ATEs and equality tests, and Appendix Figure E.3 replicates the event-study results using the binary productivity index.

Column (3) shows the effect for time spent labelling. We can infer the time employees spend on each image, as *Labelbox* tracks this time. Importantly, the timer pauses after five minutes of inactivity. Hence, if an employee is not working on their job task for a while, the timer will stop. We sum the labelling times for every workday and obtain the time employees spent labelling images for a given day. Our results suggest that there is no meaningful effect of salary delay on time spent labelling.

Table 4: Treatment Effects of Wage Withholding on Absenteeism and Work Hours

	Absent for Workday	Total Work Hours	Time Spent Labelling
ATE (One Pay Cycle Delay)	0.026	-0.008	-0.012
	(0.020)	(0.033)	(0.048)
ATE (Two Pay Cycles Delay)	0.018	0.010	-0.064
	(0.031)	(0.053)	(0.083)
ATE (Three Pay Cycles Delay)	-0.033	-0.045	-0.075
	(0.048)	(0.109)	(0.156)
Observations	14,289	12,073	12,073
Individuals	300	300	300
SE Cluster	Individual	Individual	Individual
Mean of Dep. Var.	.181	7.19	6.17

Note: This table reports the effect of delayed salary payments on worker absenteeism and effort measures, aggregated by pay cycle. The treatment effects are estimated at the worker-workday level using the Borusyak et al. (2024) imputation estimator. The daily treatment effects are combined into a single parameter spanning one pay cycle using the delta method. Standard errors are initially clustered at the individual level. This table shows the effect of salary delay on absenteeism in column (1), on total work hours in column (2) and on the total time spend labelling during the work hours in column (3).

6.2 Extensive Margin: Labor Force Participation

Empirical Strategy. We now move to the extensive margin: our goal is to identify and estimate the ATE of offering a credible salary guarantee on individuals' willingness to accept wage employment. To estimate this effect, we estimate the following specification:

$$D_i = \alpha + \lambda\, Z_{i2} + \theta\, Z_{i3} + \nu\, \text{wage}_i + \mu_t + \varepsilon_i. \tag{6}$$

Here, D_i is a binary indicator equal to one if individual i attended the orientation day and zero otherwise. $Z_{iz} = \mathbb{1}\{Z_i = z\}$ is an indicator equal to one if individual i was assigned to employment terms treatment arm $z \in \{1, 2, 3\}$, with Treatment Arm 1 (Control Arm) serving as the omitted reference category. We also include a continuous measure of the offered wage (wage_i); its coefficient ν captures the direct effect of salary amount on an individual's decision to attend the orientation day. We include date fixed effects, μ_t, to account for variation in how far the recruiting team traveled from the work location on different days, which may itself affect job take-up.

By construction of our RCT, the joint assignment of employment terms and wage is independent of potential attendance $D_i(z, w)$. This ensures that the following identifying assumption holds by design, so that θ, λ and ν are identified:

Assumption E2 (Random Treatment Assignment)**.**

$$\big(Z_i,\ \text{wage}_i\big) \perp\!\!\!\perp \big\{D_i(z, w) \colon z \in \{1, 2, 3\},\ w \in \mathcal{W}\big\}.$$

We estimate equation (6) using ordinary least squares (OLS) and a logistic regression via maximum likelihood estimation (MLE). Our primary coefficient of interest, θ, captures the causal effect of receiving a job offer corresponding to Treatment Arm 3 (Salary Certainty Arm), relative to the control group (Employment Terms Treatment Arm 1). Additionally, we use the ratio of coefficients $\frac{\theta}{\nu}$ to derive a WTP measure. Specifically, this ratio expresses the impact of receiving the high-certainty job offer relative to the impact of a salary increase, thus quantifying how large a salary increase would need to be to generate an equivalent increase in orientation-day attendance as the high-certainty employment terms. We interpret the coefficient ratio $\frac{\theta}{\nu}$ as a monetary estimate of participants' WTP for increased salary certainty.

We additionally validate our WTP estimates using the discrete-choice (conjoint) experiment conducted during the initial interview and job information session. In this experiment, participants made incentive-compatible choices among employment descriptions that systematically varied in terms of salary amount and salary certainty. Detailed methodological information are provided in Appendix K.

Results. We find that receiving a salary guarantee increases willingness to accept the job, as measured by orientation-day attendance in our in-person recruited sample. These estimates are in line with our theoretical predictions. Table 5 reports the effect of salary certainty on willingness to accept jobs among individuals who do not self-select into employment — those who were recruited in person rather than responding to the job advertisement. Column (1) and (2) show the effect of receiving a job offers corresponding to treatment arm 2 (low-certainty arm) and treatment arm 3 (high-certainty arm) relative to the control offer. Column (3) uses only the sub-sample that received an incentive to participate in the job information session while column (4) uses only the sample that participated in the job information session without incentive.

For instance, the coefficient in column (2) implies that receiving a letter outlining potential terms of employment with a salary guarantee through the automated payment system increased orientation-day attendance by 11.1 percentage points. This is a substantial effect: participation in the untreated in-person recruited sample was 44.7 percent, so the treatment effect corresponds to an effect size of approximately 25 percent.

Columns (3) and (4) provide suggestive evidence that the effect of salary certainty operates primarily through individuals who required a monetary incentive to attend the job information session. For this subsample, the coefficient on the salary guarantee remains nearly unchanged (11.8) and highly significant, with treatment effects becoming larger in magnitude given the lower baseline attendance of 40.5 percent.[61] By contrast, the effect is smaller and loses statistical significance among individuals who were willing to participate in the job information session without an incentive.[62] While we cannot rule out equality of the coefficients, this pattern suggests that some individuals are discouraged from wage employment by uncertainty around salary payments but become willing to engage in it once this uncertainty is reduced.

Table 5: Effect of Salary Guarantee ob Job Take-Up

	Logit AME	LPM	LPM	LPM
Salary Guarantee (ATE)	0.121***	0.111***	0.118***	0.078
	(0.033)	(0.032)	(0.037)	(0.068)
Salary Uncertainty (ATE)	0.005	0.002	-0.011	-0.015
	(0.038)	(0.037)	(0.043)	(0.073)
Salary (1,000 NGN)	0.005***	0.005***	0.006***	-0.001
	(0.002)	(0.001)	(0.002)	(0.003)
Observations	1,079	1,079	821	258
R-Square		0.29	0.27	0.49
Date FE	Yes	Yes	Yes	Yes
Mean of Dep. Var.	.447	.447	.405	.596
Sample	All	All	Incentivised	No Incentive

NOTE: This table reports estimates of equation (6). The results compare receiving terms of potential employment under treatment arm 2 (low-certainty arm) and treatment arm 3 (high-certainty arm) relative to the control arm (arm 1), which is the omitted category in all specifications. Standard errors are heteroskedasticity-robust. The sample is restricted to in-person recruitees. Column (1) presents average marginal effects from a logit estimation, while columns (2) to (4) report results from a linear probability model.

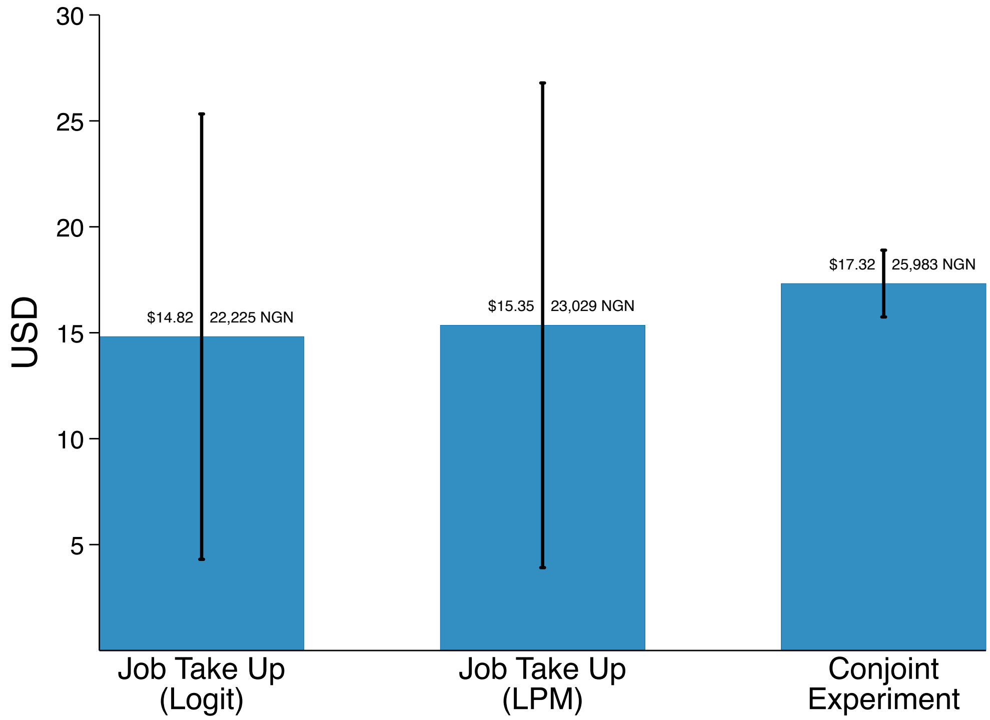
Additionally, we compare the effect of salary certainty with that of higher wages on orientation-day attendance. The coefficient on *Salary* in column (2) shows that a 1,000 NGN increase in monthly wages raises participation by 0.05 percentage points. Based on this estimate, we compute the ratio of the coefficients on *Salary Guarantee* and *Salary* to translate the effect of salary certainty into an equivalent wage increase. The first two bars in Figure 6 show

[61] The lower baseline attendance in this subsample also indicates that the treatment effects are not merely driven by the provision of a monetary incentive.

[62] Baseline participation in this subsample was higher at 59.6 percent, consistent with a greater willingness to work under current circumstances.

these ratios, corresponding to the estimates in columns (1) and (2) of Table 5.

Figure 6: WTP for higher Salary Certainty



NOTE: This figure provides estimates of the monthyl WTP for a salary guarantee. The first two bars show the ratios of the coefficients on Salary Guarantee and Salary based on the estimates in columns (1) and (2) of Table 5. The third bar reports the corresponding ratio from a conditional logit estimation (equation (34), Appendix K). Standard errors are calculated using the delta method in all cases. The conditional-logit estimate from the choice experiment (bar 3) is statistically indistinguishable from the first two estimates ($p = 0.64$ and $p = 0.74$ respectively).

We obtain an estimate of jobseekers' WTP of approximately USD 15 (about 22,500 NGN) per month.[63] The confidence intervals around the estimates are wide, however, so we cannot rule out a broader range of values. As a complementary source of evidence, we draw on the discrete choice experiment conducted during the job interviews and information sessions. In the choice experiment, jobseekers made incentive-compatible choices between hypothetical jobs that varied in salary certainty and wage levels, corresponding to the three treatment arms and analogous to the actual terms of employment. Estimating a standard conditional logit

[63]These are *per month* estimate because it is the amount of *monthly salary increase* that would yield the same effect.

model, we find a WTP of about USD 17.50 per month, remarkably close to the orientation-day attendance estimate but with much tighter confidence intervals.[64] These two approaches provide strong evidence that individuals place substantial value on salary certainty, with an implied monthly WTP between USD 15 and 18 (22,500-26,500 NGN). This corresponds to more than 100 percent of the weekly median wage and about one-third of the legal monthly minimum wage, which is widely regarded as a good salary.

6.3 Workforce Composition

Having established that salary uncertainty influences labor-force selection, we now assess whether it also affects workforce composition. Specifically, we examine whether it alters the types of workers willing to accept jobs and whether this, in turn, has productivity implications for firms. To address these questions, we use two complementary strategies. First, our main analysis uses data from the second employment round to compare the productivity of in-person recruitees hired under the guarantee with that of job-advertisement applicants who self-selected into the position. Second, we characterize the individuals induced to accept our job offer by the salary guarantee — the compliers in a standard LATE framework (Imbens and Angrist, 1994).

Estimating productivity comparisons experimentally. We examine how salary uncertainty affects workforce composition using evidence from the second employment round. In this round, both job-ad and in-person recruitees were offered contracts that included a salary guarantee. This allows us to compare individuals recruited in person — where we previously observed large selection effects from the guarantee — with jobseekers who responded to job advertisements, a group characterized by an extremely high take-up rates regardless of contract terms. To carefully assess any performance differences between the two groups that might have implications for firms' productivity, we estimate a dynamic specification comparing a range of daily productivity measures. Specifically, we estimate the following

[64] Details on the methodology and exact setup of the choice experiment are provided in Appendix K.

regression:

$$y_{itq} = \alpha + \gamma\, R_i + \mu_t + \gamma_t\, (R_i \times \mu_t) + \varepsilon_{itq}. \tag{7}$$

Here, y_{itq} denotes the outcome of interest — the continuous or binary index described in Section 6.2 — for image q, labeled by individual i on day t. Additional outcomes — absenteeism, number of correctly labelled images, work hours, and effective work time — are defined at the daily level, in which case the dependent variable becomes y_{it}. The vector μ_t represents a date fixed effects and R_i is an indicator for in-person recruitment. The coefficients γ_t trace how performance differences between the two groups evolve over time. We restrict the analysis to the first month of employment — before any individuals were exposed to salary delays — and to employees whose job offers did not specify potential salary nonpayment.

One concern with this strategy, however, is that we cannot rule out that some individuals hired from the in-person recruitment sample under high salary certainty would also have accepted a job offer without the salary guarantee. To address this concern, we implement a second strategy that explicitly identifies compliers — those induced to accept the job by the higher degree of salary certainty.

Estimation of Complier Characteristics. In this section, we again rely on our measure of job acceptance — attendance at the orientation day — to assess the characteristics of individuals induced to accept employment due to receiving an offer that included a salary guarantee. Following the terminology of the standard LATE framework (Imbens and Angrist, 1994), we distinguish three groups: always takers, who accept employment regardless of the guarantee; compliers, who accept only if the salary guarantee is offered; and never takers, who decline employment even with the salary guarantee. Relying on Abadie (2002, 2003) we can characterize the compliers using an instrumental variable framework where we use a slightly modified version of Equation (6) as the first stage. Specifically we estimate the following 2SLS framework:

$$c_i \times D_{id} = \gamma_0 + \gamma_d \widehat{D}_{id} + \delta_{w,t} + u_i \quad \text{(second stage)} \tag{8}$$

$$D_{id} = \pi_0 + \pi_1 Z_{i3} + \delta_{w,t} + e_i. \quad \text{(first stage)} \tag{9}$$

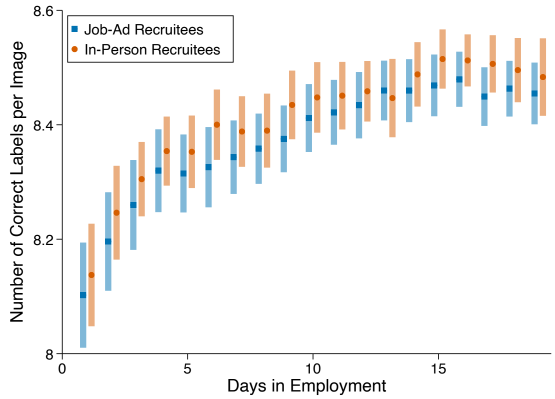
In a slight extension of notation, $D_{id} = \mathbb{1}\{D_i = d\}$ is an indicator equal to one if individual i's attendance at the orientation corresponds to $d \in \{0, 1\}$, with $d = 1$ denoting attendance and $d = 0$ non-attendance. As before, Z_{i3} denotes assignment to treatment arm 3. We nonparametrically control for date t, wage level w, and their interactions, thereby saturating the controls which is required for a LATE interpretation of 2SLS with controls (Blandhol et al., 2022). Accordingly, δ_{wt} denotes the full set of wage-by-date interaction dummies ($w \in \mathcal{W}$, $t \in \mathcal{T}$), i.e., the saturated version of the wage and μ_t controls used in (6). Let c_i denote a characteristic of individual i (Raven's score, schooling, job-search status, or WTP for salary certainty). We interact c_i with the binary attendance indicator D_{id} and regress this constructed outcome on orientation attendance, instrumented by assignment to employment terms treatment arm 3. We obtain a strong first stage with an F-statistic of 15.8, reported in Appendix Table J.1.

Under standard LATE assumptions — random assignment, exclusion, monotonicity, and relevance (discussed in detail in Appendix J) — the coefficient γ_d identifies the mean level of characteristic c among compliers with attendance status d. This allows us to characterize compliers who attended as well as those who did not attend the orientation day.[65] We pool the two complier groups and report their average characteristics, with standard errors computed using the delta method. Related applications of this approach appear, for example, in Autor and Houseman (2005) and Angrist et al. (2023), and additional details are provided in Appendix J. We use heteroskedasticity-robust standard errors.

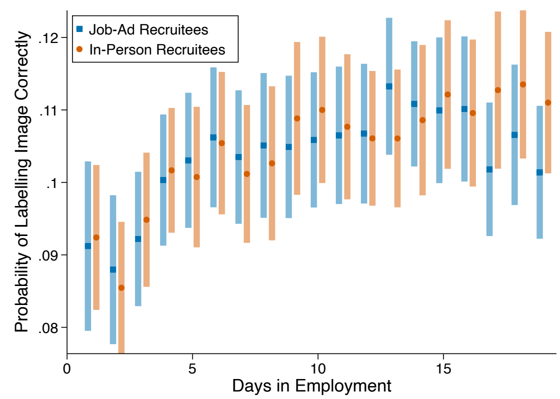
[65] Compliers who did not attend the orientation day are those who would have attended had they received a terms-of-employment offer with a salary guarantee. See Appendix J for details on the identification of this group.

Figure 7: Workforce Composition Effects

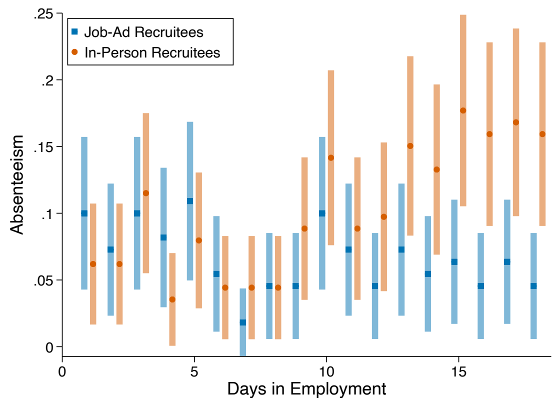
(a) Correct Labels per Image



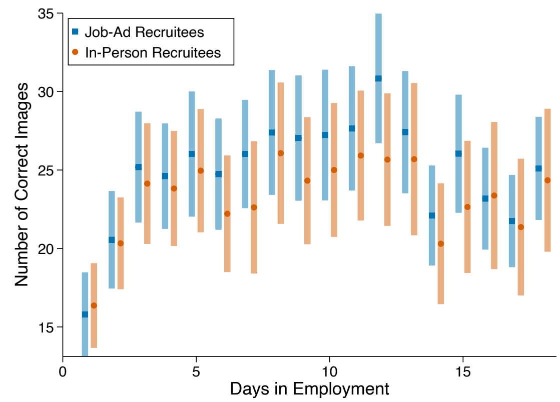
(b) Probability of Correct Label



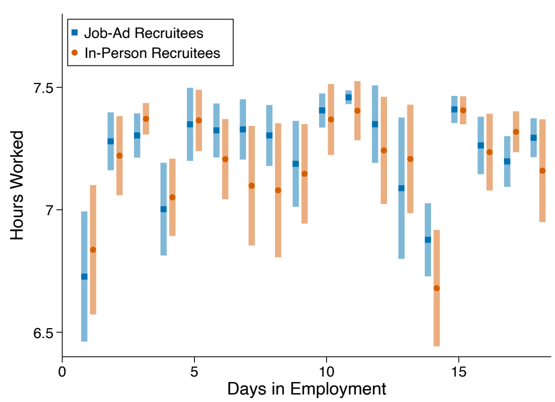
(c) Absenteeism



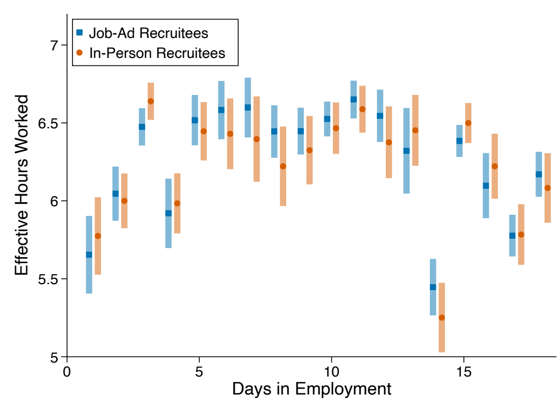
(d) Number of Correctly Labelled Images



(e) Hours Worked



(f) Effective Hours Worked



NOTE: This figure compares performance of in-person recruitees with job-ad recruitees during our second employmenr round, presenting results estimating equation (7). Panel (a) shows the number of correct labels per image, and Panel (b) displays the probability of labelling an image correctly. Both panels are estimated at the employee-image level. Panels (c), (d), (e), and (f) are estimated at the employee-workday level and report daily absenteeism, number of correctly labelled images, total hours worked, and effective hours worked, i.e. time spent labelling.

Experimental productivity Comparison Results. We find no significant differences in performance between employees recruited in-person and those recruited through job advertisements during the second employment round. Figure 7 presents results from estimating equation (7) for all six outcome measures, plotting daily averages for both groups. Panels 7a and 7b show estimates of equation (7) at the employee-image level using the continuous and binary productivity measures as outcomes. Productivity does not differ significantly on any given day, and the overall averages are also statistically indistinguishable. Appendix Section G.1 reports these comparisons of averages along with formal tests of equality between the two samples.

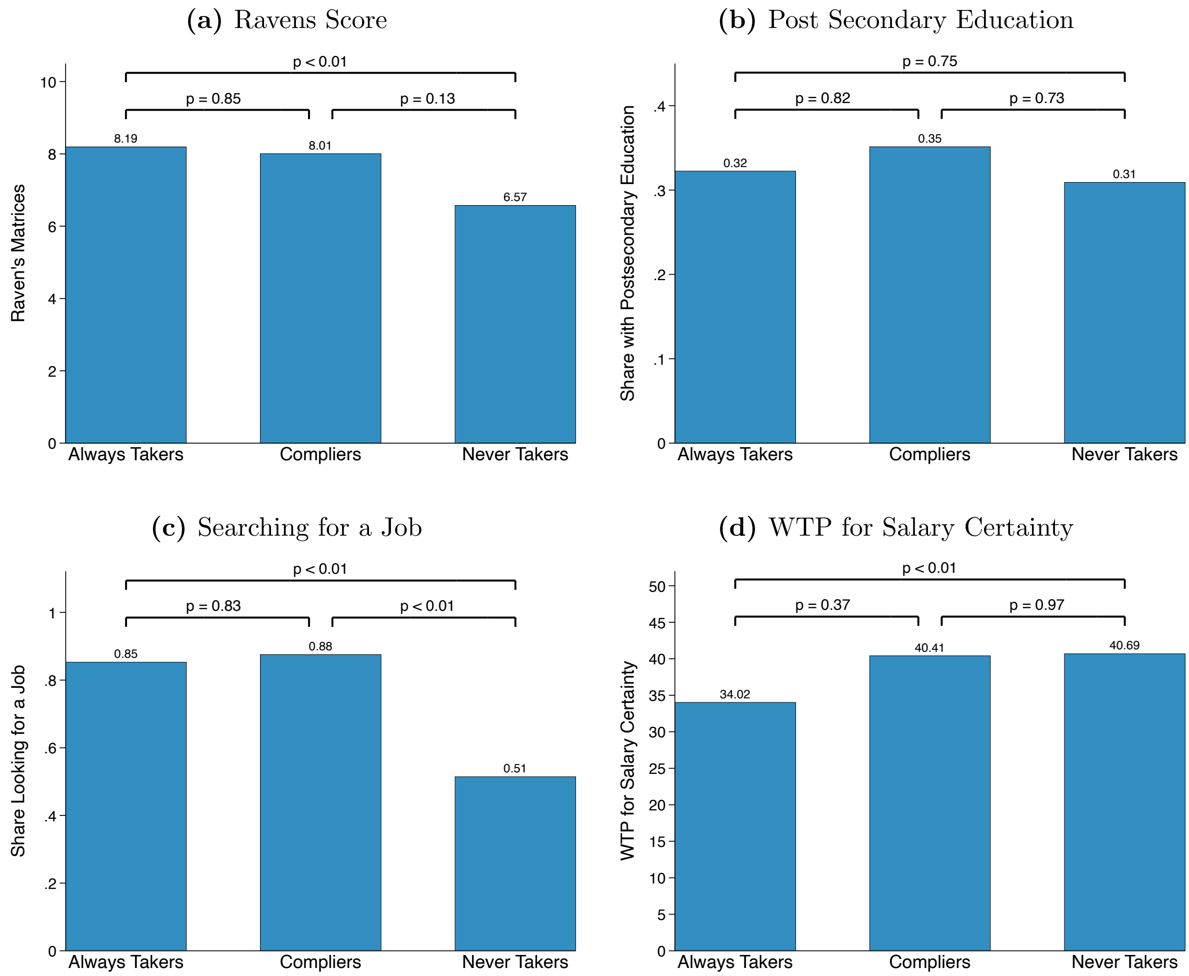
Panels 7c-7f report estimates from equation (7) at the employee-workday level. Again, both groups perform very similarly across all dimensions except absenteeism, and we cannot reject the hypothesis of equal performance. If anything, in-person recruitees are slightly more likely to be absent. The average difference in the probability to be absent is statistically significant ($p < 0.05$). Appendix Section G.2 reports overall averages and formal tests for equality of performance across the two samples.

These results imply that the performance of employees from our in-person recruitment sample is statistically indistinguishable from that of typical employees who self-selected into wage employment by responding to our job advertisements. Hence, firms would not be expected to gain significantly in productivity from hiring these additional workers who are only willing to work when salary certainty is high, assuming firms are not labor constrained.

Complier Characteristics Results. Compliers are much more similar to always-takers than to never-takers on observable characteristics, as shown in Figure 8. We report results for three key characteristics: Raven's score, attainment of post-secondary education, and job search status. However, compliers are much more similar to never takers in their WTP for salary certainty, which we use to proxy risk preferences.

Across the three observable characteristics, compliers are statistically indistinguishable from

Figure 8: Complier Characteristics



Note: This figure reports results from equation (9), in which we instrument attendance at the orientation day, D_i, with assignment to the high-certainty treatment arm. We report characteristics of the complier population (i.e., individuals induced to accept the job because of higher salary certainty), always-takers (i.e., those who accept the job even without salary certainty), and never-takers (i.e., those who do not accept the job even when higher salary certainty is offered). Panel (a) shows results for the Raven's score, Panel (b) for schooling, and Panel (c) for job-search status. These panels suggest that compliers are more similar to always-takers on observable characteristics. Panel (d) reports WTP for higher salary certainty and suggests that compliers are more similar to never-takers in their risk preferences.

always-takers, in contrast to never-takers, who differ noticeably across several dimensions. Panel (a) shows that compliers' Raven's scores are nearly identical to those of always-takers and somewhat higher than those of never-takers, though the latter difference is not statistically significant. In Panel (b), post-secondary education attainment is similar across all three groups, with no significant differences. Panel (c) reveals a stark contrast in job search status: compliers, like always-takers, are highly likely to report active job search, whereas never-takers are much less likely. However, Panel (d) indicates that compliers' WTP for salary certainty is much more similar to that of never-takers. Even though the difference between always-takers and compliers is not statistically significant, the evidence suggests that compliers are willing to pay more for salary certainty, consistent with a greater degree of risk aversion. This pattern aligns with the third prediction of our theoretical framework: selection could be driven by higher ability or greater risk aversion. We find no evidence for the former and suggestive evidence for the latter.

7 Conclusion

This paper studies the prevalence and consequences of overdue and unpaid salary payments in Nigeria, a practice that likely affects the welfare of millions of workers. Our results offer a new perspective for understanding this phenomenon. The theoretical framework illustrates that in settings such as Nigeria — where liquidity shocks are believed to be common and formal contract enforcement is weak — workers may have an incentive to increase effort when wages are unpaid because the expected value of future payment rises. The experimental evidence is consistent with this prediction: we find that workers' effort increases when their wages are delayed. We also find that signaling salary reliability increases job take-up but does not attract more productive workers.

Taken together, our findings suggest that firms face minimal productivity implications from engaging in wage withholding, provided that labor markets remain slack and firms are not labor constrained. To gauge the potential magnitude, we combine our intensive- and extensive-margin estimates in a back-of-the-envelope calculation. The exercise compares observed productivity with a counterfactual in which firms do not withhold wages, accounting for (i) the small effort gains from delayed pay, (ii) the productivity of workers who would enter employment if wages were guaranteed, and (iii) the fact that salary reliability increases job take-up by 25 percent. Even under conservative assumptions — using the lower bound of the first effect, the upper bound of the second, and assuming firms could replace 25 percent of their workforce with 'new' workers who would not otherwise accept wage employment — the resulting productivity loss amounts to only about 0.2 percent.[66] This negligible effect helps explain why wage withholding can persist: firms have little incentive to refrain from the practice.

Nevertheless, our results suggest that wage withholding is socially inefficient due to its large costs for workers and their willingness to forgo a substantial share of their salary for payment certainty. Even under the most favorable assumptions, the productivity gains from delayed pay seem too small to justify the practice. Firms could instead reduce base wages slightly, introduce simple performance-based incentives, and pay on time — lowering their total wage bill while achieving the same effort increase. For instance, the upper bound of the confidence interval from the productivity estimates (column (2) of Table 3) suggests that effort could rise by up to 12 percent when wages are delayed. The same effort gains could likely be replicated through a bonus of roughly 10 NGN (0.6 US cents) per correctly labeled image, corresponding to about 5,600 NGN (USD 3.75) per worker per month in our setting.[67] This amount is far below workers' own valuation of reliable salary payments, which we estimate at around 22,500 NGN (USD 15) per month. Thus, wage withholding persists not because it is efficient, but because firms face little consequences from engaging in this practice.

[66] See Appendix H.1 for additional details on this calculation.
[67] See Appendix H.2 for additional details on this calculation.

References

Abadie, Alberto, "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 2002, *97* (457), 284–292.

— , "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 2003, *113* (2), 231–263.

Africanews, "Street sweepers in Abidjan protest over unpaid wages," August 2024. Accessed: November 1, 2025. Available at: https://www.africanews.com/2016/02/16/street-sweepers-in-abidjan-protest-over-unpaid-wages/.

Akerlof, George A. and Janet L. Yellen, "The fair wage-effort hypothesis and unemployment," *The Quarterly Journal of Economics*, 1990, *105* (2), 255–283.

Akomolafe, Mohammed Akinola, "The Rule of Law and its Practice in Nigeria: An Assessment," *Makurdi Owl Journal of Philosophy (MAJOP)*, 2021, *2* (2), 13–28.

Akuopha, Ochuko, "Unpaid Salaries: Pipeline Surveillance Shut Down Oil Facilities in Delta," April 2023. Vanguard (Ughelli), accessed October 17, 2025.

Alfandari, Gilles and Mark Edwin Schaffer, ""Arrears" in the Russian Enterprise Sector," in Simon Commander, Qimiao Fan, and Mark Edwin Schaffer, eds., *Enterprise Restructuring and Economic Policy in Russia*, Washington, DC: World Bank, 1996, pp. 87–139.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

Angrist, Joshua, Peter Hull, and Christopher Walters, "Chapter 1 - Methods for measuring school effectiveness," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 7, Elsevier, 2023, pp. 1–60.

Animbom, Sah Terence, Fomusoh Wadyeh Rachel, and Ndong Clinton Toh, "Unpaid wages, paid press conferences: How politicians manipulate the news in Cameroon," *Unbias The News*, June 2022. Accessed: November 1, 2025. Available at: https://unbiasthenews.org/unpaid-wages-cameroon-journalism/.

Ashraf, Nava, Oriana Bandiera, and Katharina Jack, "No Margin, No Mission? A Field Experiment on Incentive Pay for Teachers in Zambia," *Journal of Public Economics*, 2014, *120*, 60–71.

— , — , Sanghamitra Lee, and Emily Nix, "Togethering to Win: Experimental Evidence on Team Incentives in Zambia," *American Economic Journal: Applied Economics*, 2018, *10* (2), 1–35.

Autor, David H. and Susan N. Houseman, "Do Temporary Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from Random Assignments," Staff

Working Paper 05-124, W.E. Upjohn Institute for Employment Research October 2005. Revised from January 2005.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Incentives for Managers and Inequality among Workers: Evidence from a Firm-level Experiment," *Quarterly Journal of Economics*, 2007, *122* (2), 729–773.

Barker, Nathan, Inbar Amit, Alison Andrew, Robert Garlick, Kate Orkin, and Carol Nekesa, "Flexibility versus Performance: The Determinants of Labor Contracts in Nairobi, Kenya," November 2024. Accessed: July 12, 2025. Available at: https://www.nathan-barker.com/_files/ugd/6e7d59_072f0fddd7154d0aa8fa8697db311c45.pdf.

BBC Sport, "Delayed wages, fan attacks and no contracts: why Africa's young soccer stars keep leaving," 2019. Accessed: November 1, 2025. Available at: https://www.bbc.com/sport/football/48894725.

Becker, Gary S. and George J. Stigler, "Law Enforcement, Malfeasance, and Compensation of Enforcers," *The Journal of Legal Studies*, 1974, *3* (1), 1–18.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 09 2006, *93* (3), 491–507.

Bernhardt, Annette, Ruth Milkman, Nik Theodore, Douglas Heckathorn, Mirabai Auer, James DeFilippis, Ana Luz González, Victor Narro, Jason Perelshteyn, Diana Polson, and Michael Spiller, "Broken Laws, Unprotected Workers: Violations of Employment and Labor Laws in America's Cities," Technical Report, National Employment Law Project, UCLA Institute for Research on Labor and Employment 2009.

Bewley, Truman F, "Why not cut pay?," *European Economic Review*, 1998, *42* (3), 459–490.

Bewley, Truman F., *Why Wages Don't Fall During a Recession*, Cambridge, MA: Harvard University Press, 1999.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, "When is TSLS Actually LATE?," Working Paper 29709, National Bureau of Economic Research January 2022.

Blattman, Christopher and Stefan Dercon, "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 2018, *10*, 1–38.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *The Review of Economic Studies*, 02 2024, *91* (6), 3253–3285.

Boyarchuk, Dmytro, Lilia Maliar, and Serguei Maliar, "The consumption and welfare implications of wage arrears in transition economies," *Journal of Comparative Economics*, 2005, *33* (3), 540–564.

Breza, Emily and Supreet Kaur, "Labor Markets in Developing Countries," Working Paper 33908, National Bureau of Economic Research June 2025.

_ , _ , and Yogita Shamdasani, "The Morale Effects of Pay Inequality," *Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Buehren, Niklas, Virginia Ceretti, Ervin Dervisevic, Markus Goldstein, Leora Klapper, Tricia Koroknay-Palicz, and Simone Schaner, "Salary Delays and Overdrafts in Rural Ghana," *AEA Papers and Proceedings*, May 2018, *108*, 449–52.

Carranza, Eliana, Aletheia Donald, Florian Grosset, and Supreet Kaur, "The Social Tax: Redistributive Pressure and Labor Supply," NBER Working Paper 30438, National Bureau of Economic Research 2022.

Cassan, Guilhem, Daniel Keniston, and Tatjana Kleineberg, "A Division of Laborers: Identity and Efficiency in India," NBER Working Paper 28462, National Bureau of Economic Research 2021.

Cefala, Lauren, Supreet Kaur, Heather Schofield, and Yogita Shamdasani, "Habit Formation in Labor Supply," Technical Report 2024. Yale University. *mimeo*. Date Accessed: November 4, 2025. Available at: https://economics.yale.edu/sites/default/files/2024-11/CKSS_draft.pdf.

Clarke, Simon, "Trade unions and the non-payment of wages in Russia," *International Journal of Manpower*, 1998, *19* (1/2), 68–94.

Clemens, Jeffrey and Michael R. Strain, "Understanding "Wage Theft": Evasion and avoidance responses to minimum wage increases," *Labour Economics*, 2022, *79*, 102285.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey," *The Econometrics Journal*, 06 2022, *26* (3), C1–C30.

Donald, Alexia and Florentin Grosset, "Complementarities in Labor Supply: Joint Commute and Work Decisions," *Working Paper*, 2024.

Dunn, Abe, Joshua D. Gottlieb, Adam Hale Shapiro, Daniel J. Sonnenstuhl, and Pietro Tebaldi, "A Denial a Day Keeps the Doctor Away," *Quarterly Journal of Economics*, February 2024, *139* (1).

Earle, John S. and Klara Sabirianova Peter, "Complementarity and Custom in Wage Contract Violation," *The Review of Economics and Statistics*, November 2009, *91* (4), 832–849.

_ and Klara Z. Sabirianova, "How Late to Pay? Understanding Wage Arrears in Russia," *Journal of Labor Economics*, July 2002, *20* (3), 661–707.

Falk, Armin, "Gift Exchange in the Field," *Econometrica*, 2007, *75* (5), 1501–1511.

Fernandes-Brough, Jacinta, "Nigeria Women's World Cup Players Speak Out Over Unpaid Wages," September 2023. Accessed: October 17, 2025. Available at: https://theathletic.com/4898797/2023/09/30/nigeria-womens-world-cup-wages/.

Flynn, Suzanne and Mario Pessoa, "Prevention and Management of Government Expenditure Arrears," Technical Report TNM/14/03, International Monetary Fund, Fiscal Affairs Department, Washington, DC 2014.

Freeman, Richard B., Xiaofei Pan, Xiaolan Yang, and Maoliang Ye, "Team incentives and lower ability workers: A real-effort experiment," *Journal of Economic Behavior & Organization*, 2025, *233*, 106986.

Galvin, Daniel J., "Deterring Wage Theft: Alt-Labor, State Politics, and the Policy Determinants of Minimum Wage Compliance," *Perspectives on Politics*, 2016, *14* (2), 324–350.

Gausi, Tamara, "“What should I do? Stay at home and starve?” – wage theft leaves Zimbabwe's workers with an impossible dilemma," *Equal Times*, Dec 2018. Accessed: November 1, 2025. Available at: https://www.equaltimes.org/what-should-i-do-stay-at-home-and.

Gerry, Christopher J., Carmen A. Li, and Byung-Yeon Kim, "The Gender Wage Gap and Wage Arrears in Russia: Evidence from the RLMS," *Journal of Population Economics*, 2004, *17* (2), 267–288.

Guiteras, Raymond P. and B. Kelsey Jack, "Productivity in piece-rate labor markets: Evidence from rural Malawi," *Journal of Development Economics*, 2018, *131*, 42–61.

Huck, Steffen, Andrew J. Seltzer, and Brian Wallace, "Deferred Compensation in Multiperiod Labor Contracts: An Experimental Test of Lazear's Model," *American Economic Review*, 2011, *101* (2), 819–843.

Igwe, Isaac O. C., "Rule of Law and Constitutionalism in Nigerian Democracy: A Critical Relativism Discussion in the Context of International Law," *Athens Journal of Law*, July 2021, *7* (3), 317–334.

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

IndustriALL Global Union, "Senegal's garment workers demand end to exploitation at Sartorisen," July 2025. Accessed: November 1, 2025. Available at: https://www.industriall-union.org/senegals-garment-workers-demand-end-to-exploitation-at-sartorisen.

Jeffay, John, "Diamond Mine Workers Strike over Unpaid Salaries," Oct 2024. Accessed: November 1, 2025. Available at: https://www.idexonline.com/FullArticle?Id=49885.

Kapoor, Varun, "Wage theft, reneging and liquidity constraints in informal labor markets," 2025. Harvard University. *mimeo.* Date Accessed: November 4, 2025. Available at: https://uchicago.app.box.com/s/5w6uaw1h3sqz63hd11o5ja26mh8wcqdf.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan, "Self-Control at Work," *Journal of Political Economy*, 2015, *123* (6), 1227–1277.

__ , **Sendhil Mullainathan, Suanna Oh, and Frank Schilbach**, "Do Financial Concerns Make Workers Less Productive?," *The Quarterly Journal of Economics*, 2 2025, *140* (1), 635–689.

Krueger, Alan B. and Alexandre Mas, "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires," *Journal of Political Economy*, 2004, *112* (2), 253–289.

Langener, AM, AW Kramer, W van den Bos, and HM Huizenga, "A shortened version of Raven's standard progressive matrices for children and adolescents," *British Journal of Developmental Psychology*, March 2022, *40* (1), 35–45. Epub 2021 May 27.

Lazear, Edward P., "Why is there mandatory retirement?," *Journal of Political Economy*, 1979, *87* (6), 1261–1284.

__ , "Performance Pay and Productivity," *American Economic Review*, 2000, *90* (5), 1346–1361.

Lehmann, Hartmut and Jonathan Wadsworth, "Wage Arrears and Inequality in the Distribution of Pay: Lessons from Russia," in Solomon W. Polachek and Olivier Bargain, eds., *Aspects of Worker Well-Being*, Vol. 26 of *Research in Labor Economics*, Leeds: Emerald Group Publishing Limited, 2007, pp. 125–155.

__ , __ , **and Alessandro Acquisti**, "Grime and Punishment: Job Insecurity and Wage Arrears in the Russian Federation," *Journal of Comparative Economics*, 1999, *27* (4), 595–617.

Liu, Licheng, Ye Wang, and Yiqing Xu, "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data," *American Journal of Political Science*, 2024, *68* (1), 160–176.

Mas, Alexandre, "Pay, Reference Points, and Police Performance," *Quarterly Journal of Economics*, 2006, *121* (3), 783–821.

Milkman, Ruth, Ana Luz González, Victor Narro, Annette Bernhardt, Nik Theodore, Douglas Heckathorn, Mirabai Auer, James DeFilippis, Jason Perelshteyn, Diana Polson, and Michael Spiller, "Wage Theft and Workplace Violations in Los Angeles: The Failure of Employment and Labor Law for Low-Wage Workers," Technical Report, UCLA Institute for Research on Labor and Employment, Los Angeles, CA 2010.

National Bureau of Statistics, "Nigeria Labour Force Survey (NLFS) Report, Q2 2024," November 2024. Accessed: November 1, 2025. Available at: https://www.nigerianstat.gov.ng/elibrary/read/1241594.

News Ghana, "Nigerian Doctors Launch Strike Over Unpaid Allowances Stretching Back a Decade," November 2025. Accessed: November 1, 2025. Available at: https://www.newsghana.com.gh/nigerian-doctors-launch-strike-over-unpaid-allowances-stretching-back-decade/.

— , "Unpaid Nurses Storm Accra Over 10-Month Salary Arrears," October 2025. Accessed: November 1, 2025. Available at: https://www.newsghana.com.gh/unpaid-nurses-storm-accra-over-10-month-salary-arrears/.

Oh, Suanna, "Does Identity Affect Labor Supply?," *American Economic Review*, 2023, *113* (8), 2055–2083.

Punch, "Oil Workers Protest Over Unpaid Salaries," 2023. Accessed: October 17, 2025. Available at: https://punchng.com/oil-workers-protest-over-unpaid-salaries/.

PwC, "MSME Survey 2020 Final," Survey Report, PricewaterhouseCoopers Nigeria 2020. Accessed: July 2, 2025. Available at: https://www.pwc.com/ng/en/assets/pdf/pwc-msme-survey-2020-final.pdf.

— , "MSME Survey 2024: Building resilience – Strategies for MSME success in a changing landscape," Survey Report, PricewaterhouseCoopers Nigeria 2024. Accessed: July 12, 2025. Available at: https://www.pwc.com/ng/en/assets/pdf/2024-pwc-msme-survey-report.pdf.

Radio Tamazuj, "Sudan: STUF lodges complaint with ILO over unpaid salaries," August 2023. Accessed: November 1, 2025. Available at: https://www.radiotamazuj.org/en/news/article/sudan-stuf-lodges-complaint-with-ilo-over-unpaid-salaries.

Robinson, Erin, Ha T. Nguyen, Scott Isom, Sara A. Quandt, Joseph G. Grzywacz, Haiying Chen, and Thomas A. Arcury, "Wages, Wage Violations, and Pesticide Safety Experienced by Migrant Farmworkers in North Carolina," *NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy*, 2011, *21* (2), 251–268.

Roth, Jonathan and Pedro H. C. Sant'Anna, "Efficient Estimation for Staggered Rollout Designs," *Journal of Political Economy Microeconomics*, 2023, *1* (4), 669–709.

— **, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe**, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Shapiro, Carl and Joseph E. Stiglitz, "Equilibrium unemployment as a worker discipline device," *The American Economic Review*, 1984, *74* (3), 433–444.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199. Themed Issue: Treatment Effect 1.

Sánchez de la Sierra, Raúl, "Whither Formal Contracts?," *Econometrica*, 2021, *89* (5), 2341–2373.

— **, Kristof Titeca, Haoyang (Stan) Xie, Aimable Amani Lameke, and Albert Malukisa Nkuku**, "The Real State: Inside the Congo's Traffic Police Agency," *American Economic Review*, December 2024, *114* (12), 3976–4014.

Train, Kenneth E., *Discrete Choice Methods with Simulation*, Cambridge, UK: Cambridge University Press, 2003.

US Department of State, Bureau of Democracy, Human Rights and Labor, "Country Reports on Human Rights Practices for 2023: Nigeria," Human Rights Report, United States Department of State 2023. Accessed: July 12, 2025. Available at: https://www.state.gov/wp-content/uploads/2024/02/528267_NIGERIA-2023-HUMAN-RIGHTS-REPORT.pdf.

Vazquez-Bare, Gonzalo, "Identification and estimation of spillover effects in randomized experiments," *Journal of Econometrics*, 2023, *237* (1), 105237.

Appendices

Contents

A Descriptive Survey

A.1 Methodology and Data

We collected data in Lagos between June and August 2025. A team of seven enumerators surveyed 1,279 employees and former employees (for instance individuals who were self-employed but worked as employees for someone else before) across a range of market settings.[68] In addition, enumerators visited industrial areas and office complexes to reach workers employed in larger firms. Surveys were administered on tablet computers, and participants were reimbursed with 1,000 NGN (approximately \$0.66). The average survey lasted 21.2 minutes, with a median duration of 21.8 minutes.[69] Enumerators were assigned to predetermined areas, which they subdivided into sections to minimize overlap and reduce the risk of interviewing the same individual more than once. Each enumerator began at a different location within their section, and the tablet software then instructed them to skip a randomized number of individuals before approaching the next potential participant. Appendix Table A.1 summarizes the characteristics of the respondents. The average participant was 30.8 years old, and 50 percent were female. 32 percent reported actively searching for a job, and the mean monthly income was approximately 199,000 NGN (about \$132). The sample includes both employed (42 percent) and self-employed individuals (56 percent), and spans a wide range of firm sizes. Respondents were also relatively well-educated, mirroring the patterns observed in our recruitee samples.

[68] One of the co-authors also conducted a small subset of surveys.

[69] For 16 surveys, the recorded duration was an implausible 2.18 minutes, with all 16 also showing exactly the same value — an outcome that is statistically highly unlikely. We therefore attribute this to a software error in the *ODK* start and end functions and exclude these cases when calculating the mean and median durations.

Table A.1: Characteristics of the Survey Participants

	Mean	SD	10th Percentile	90th Percentile	Observations
Characteristics					
Age	30.81	8.08	22.00	41.00	1,279
% Female	0.50	0.50	0.00	1.00	1,279
% Searching for a Job	0.32	0.46	0.00	1.00	1,279
Income	199,131.32	239,910.47	40,000.00	500,000.00	1,183
Employment Status					
Unemployed	0.02	0.15	0.00	0.00	1,279
Self-Employed	0.42	0.49	0.00	1.00	1,279
Employed	0.56	0.50	0.00	1.00	1,279
At small firm (<10 employees)	0.46	0.50	0.00	1.00	714
At medium firm (10 to 50 employees)	0.30	0.46	0.00	1.00	714
At large firm (>50 employees)	0.24	0.43	0.00	1.00	714
Highest Schooling					
Secondary School	0.53	0.50	0.00	1.00	1,279
Vocational Training	0.00	0.03	0.00	0.00	1,279
National Diploma	0.23	0.42	0.00	1.00	1,279
University	0.23	0.42	0.00	1.00	1,279

NOTE: This Table shows the characteristics of the survey participants. Individuals had the option not to answer questions and several chose not to answer questions about their income, which explains the lower number of observations. To illustrate employment at different firm types, we subset the sample to individuals who indicated they were employed.

A.2 Additional Descriptive Results

In this section, we provide additional descriptive results.

Concerns about large and small firms. We complement Figure 1e with supplemental results. The results are presented in Table A.2. We investigate individuals' concerns for working at different types of firms. We distinguish between large and small firms. The exact wording used in the survey for large firms is "Imagine you work for a large and established company with a written work agreement. Would you be concerned about."

For small firms, we wanted to distinguish between firms owned by a friend or family member, and by someone not necessarily from the respondent's social network. Finding employment through social networks — particularly through family connections — is very common in the context of our study.

Column (4) of Table A.2 compares concern levels between family-owned and non-family-owned small firms. Employment in a family firm reduces the share of respondents expressing any concern about the hypothetical job, and this difference is statistically significant. However, for most specific concerns, the differences between family and non-family ownership are not statistically significant. Finally, we compare concern levels between non-family-owned small firms and large firms. The hypothetical scenario of working for a large firm reduces participants' concerns across almost all dimensions, and these differences are statistically significant.

Table A.2: Firm Concerns

	Mean Response			Difference	
	Large Firm	Small Firm (non-family owned)	Small Firm (family owned	Small firms family - non-family owned	Large Firm-small firm
Concern: being fired and without job	0.20	0.18	0.15	0.03	0.02
				(0.021)	(0.021)
Concern: paid as agreed and on time	0.31	0.50	0.45	0.04	-0.19***
				(0.028)	(0.024)
Concern: employer sticking to work agreement	0.21	0.27	0.33	-0.05**	-0.07***
				(0.026)	(0.020)
Concern: poorly treated b/c of ethnicity	0.03	0.05	0.02	0.03**	-0.01
				(0.011)	(0.009)
Concern: poorly treated b/c of religion	0.03	0.05	0.02	0.03**	-0.02**
				(0.010)	(0.007)
No Concern	0.40	0.26	0.31	-0.06**	0.15***
				(0.025)	(0.022)
Observations	1,279	654	625	1,279	654

Note: This table presents individuals' concerns about working at different types of firms and replicates and extends the results shown in Panel (e) of Figure 1. We report concerns for large firms (column 1) and for small informal firms, distinguishing between those owned by family members (column 3) and those not owned by family members (column 2). Finding employment through social networks — particularly family — is very common in the context of our study. Column 4 shows differences in concern levels between family-owned and non-family-owned small firms. Employment at a family firm reduces the share of respondents expressing any concern about the hypothetical job, and this difference is statistically significant. However, for most specific concerns, the difference between family and non-family ownership is not tatistically significant. Finally, we compare concern levels between non-family-owned small firms and large firms. The hypothetical scenario of working for a large firm reduces participants' concerns across almost all dimensions, and these differences are statistically significant.

Workers' experiences. We provide additional results around workers' experiences with unpaid or delayed salaries. We document how workers usually coped with such a situation by asking about coping strategies with unpaid or delayed salaries. We additionally ask about workers experiences withing the firm, if they are the only person who were affected by delayed or unpaid salaries, if multiple people were affected, or if all people were affected. The results are presented in Table A.3.

Table A.3: Coping with Salary Non-Payment and Its Extent within the Firm

	Mean	SD	Observations
Panel A: Coping with Delayed Salaries			
Borrowed from family/friends	0.31	0.46	383
Borrowed from moneylender	0.04	0.20	383
Had Savings	0.45	0.50	383
Had other income	0.33	0.47	383
Panel B: Wage Withholding Among Employees			
Participant was the only unpaid employee	0.23	0.42	383
Some employees were unpaid	0.46	0.50	383
All employees were unpaid	0.31	0.46	383

NOTE: This table only includes the subset of descriptive survey respondents who have experienced salary delays or nonpayment. Panel A reports how respondents coped during these periods. Respondents could select multiple coping strategies from the four options shown. Panel B reports the extent to which employers failed to pay their employees during these episodes. Responses reflect a single-choice question. The table shows that among those who experienced salary delay or nonpayment, different degrees of nonpayment or delay by the employer — whether a single employee, some employees, or all employees were affected — are relatively common.

Wage withholding, social norms, and social contexts. We document two additional aspects related to wage withholding and social norms and social contexts in Table A.4. Panel A presents a measure of social norms surrounding wage withholding. One interpretation of the prevalence of this practice is that it may be socially accepted, giving firms implicit legitimacy to withhold their employees' wages. We show that this is not the case: about 70 percent of workers stated that firms should pay their employees on time. Second, we complement this measure of social norms with evidence on prevalence within workers' social networks in Panel B. When asked whether they personally knew someone who had experienced delayed or unpaid salaries, 49 percent responded affirmatively.

Table A.4: Wage Withholding in the Context of Social Norms and Social Networks

	Mean	SD	Observations
Panel A: Social Norm of Wage Payments			
Employers should pay their workers	0.70	0.46	1,279
Employers may delay payments if in trouble	0.22	0.42	1,279
Workers should temporarily forgo wages if employer is in trouble	0.07	0.26	1,279
Panel B: Wage Withholding in Social Network			
Knows someone who experienced salary difficulties	0.49	0.50	1,279

Note: This table reports participants' perceptions of social norms around wage withholding and the extent to which wage withholding affects their social networks. Panel A presents the measure of social norms. Participants were asked to select one of three statements describing what they consider the most acceptable behavior; the table shows the share of respondents who agreed with each statement. Panel B reports how many respondents know someone in their social network who has experienced some form of salary difficulty (i.e., payment delays, and partial or complete nonpayment). The small differences in sample size occur because "don't know" was a valid response option, and observations with this answer are excluded.

Perceived Risk of Experiencing Wage Withholding. We ask survey participants about the likelihood that they would experience salary difficulties — defined as delayed or unpaid wages — if they were to start working at a firm tomorrow. Respondents were asked about four different firm types: small informal firms, large formal firms, a new firm (which, by definition, has no reputation), and *Dangote*, a large and well-known firm with a strong reputation for paying employees on time. For each firm type, participants answered a simple yes-or-no question indicating whether they expected to experience salary difficulties.

Table A.5: Additional Beliefs about Salary Payments

	Mean	SD	Observations
Perceived Risk of Salary Difficulty			
At new firm	0.44	0.28	1,271
At reputable firm (Dangote)	0.20	0.21	1,279
At small firm	5.20	2.49	1,279
At large firm	3.10	2.43	1,277

Note: This table reports participants' beliefs about the likelihood of experiencing salary difficulties (i.e., payment delays, and partial or complete nonpayment) across different firm types. Respondents were asked about small informal firms, large formal firms, a new firm (which, by definition, has no reputation), and *Dangote* — a large and well-known firm with a good reputation of paying employees. The slightly different number of observations is due to some participants responding "don't know" to particular questions.

Preferences for Self-Employment. We ask survey participants whether they would prefer self-employment — in its typical local form, i.e., small-scale, subsistence entrepreneurship — or wage employment. We find that the vast majority, 82 percent, state a preference for self-employment over wage employment. We then ask about the underlying reasons, allowing for multiple responses to obtain an accurate picture of people's reasoning. A main concern relates to the perceived risk of wage employment: 33 percent report that one reason they prefer self-employment is that it provides a more stable and secure source of income compared to being employed. This likely reflects the uncertainty associated with wage employment due to the prevalence of delayed or unpaid wages. Results are presented in Table A.6.

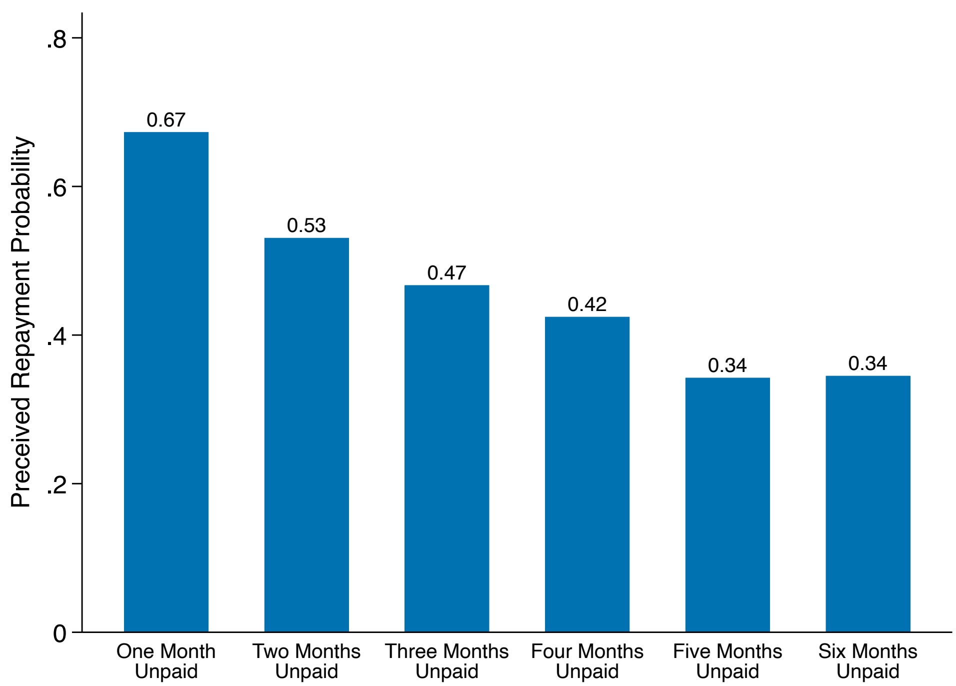
Table A.6: Preferences for Self-Employment

	Mean	SD	10th Percentile	90th Percentile	Observations
Preference for Self-Employment	0.82	0.38	0.00	1.00	1,279
Reasons for Self-Employment Preference					
Pref. for self-empl: can grow business and make more money	0.47	0.50	0.00	1.00	1,049
Pref. for self-empl: want to be one's own boss	0.39	0.49	0.00	1.00	1,049
Pref. for self-empl: more stable income than employemnt	0.33	0.47	0.00	1.00	1,049
Pref. for self-empl: Concerned about employer sticking to work agreement	0.07	0.26	0.00	0.00	1,049
Pref. for self-empl: Concerned to be fired and without job	0.09	0.28	0.00	0.00	1,049
Pref. for self-empl: employment locks in, good opportunity can be missed	0.09	0.28	0.00	0.00	1,049
Pref. for self-empl: self-employment is status symbol	0.07	0.25	0.00	0.00	1,049

NOTE: This table shows that the vast majority of survey respondents of our survey (82%) prefer self-employment over wage employment. Participants were asked whether they would rather be self-employed or wage-employed. For those who preferred self-employment, we followed up by asking for their reasons and allowed multiple responses (this reflects the different sample size). The most frequently cited reason (47%) is the belief that they can grow their own business and earn more than in wage employment. Remarkably, 33% list greater stability and security as a main reason for preferring self-employment, reflecting the uncertainty created by widespread wage withholding.

Beliefs about Repayment of Arrears. To understand people's beliefs about the repayment of outstanding wage balances, we asked survey participants to assess how likely they thought it was that such balances would eventually be repaid in a series of hypothetical scenarios. Participants were first asked to consider a situation in which they had been unpaid for one month and to rate, on a scale from 0 to 10, how likely they were to receive both future salaries and any outstanding arrears. Each participant was then randomly assigned one additional scenario, in which they would be unpaid for 2, 3, 4, 5, or 6 months, and was asked the same question again. The numeric responses (0-10) were subsequently converted into probabilities for analysis.

Figure A.1: Beliefs about the Likelihood of Payment after Sustained Nonpayment



NOTE: This figure illustrates respondents' beliefs about the likelihood of receiving salary payments after a sustained period of non-payment. Survey participants were first asked to consider a hypothetical situation in which they had been unpaid for one month and to rate, on a scale from 0 to 10, how likely they were to receive future salaries and any outstanding balances. Here, 0 indicated being certain that no payment would be received and 10 indicated being certain that full payment would be received. Each participant was then randomly assigned one additional scenario, in which they would be unpaid for 2, 3, 4, 5, or 6 months, and was asked the same question again. The numeric responses (0-10) were subsequently converted into probabilities.

B Applicants

B.1 In-Person Recruitment

In-person recruitment was carried out by a team of eight enumerators. For the recruitment task, they were formally affiliated with the recruitment agency *Unlocking Creativity* and wore clothing and ID cards displaying the agency's branding.

Enumerators recruited participants in pre-selected markets and public areas. Each was equipped with a tablet computer running *ODK Collect* to conduct the survey. The survey form implemented a pre-specified randomization protocol. Before approaching an individual, enumerators followed a skip pattern that required them to pass over a randomly determined number of suitable persons (between one and ten). The software also randomly specified whether to approach a man or a woman (each with 50 percent probability) and, independently, whether to approach someone who appeared to be self-employed with a small business (60 percent probability) or someone who was not self-employed and appeared to be passing by (40 percent probability). Appendix Figure B.4 illustrates the number of interactions of the enumerators per recruitment day.

Table B.7: Balance of Employee Baseline Sample

	Treatment Arm 1		Treatment Arm 2		
	Mean	SD	Mean	SD	Difference
Characteristics					
Age	29.48	10.21	31.31	9.76	-1.82
					(1.868)
% Female	0.55	0.51	0.54	0.50	0.01
					(0.093)
% Searching for a Job	1.00	0.00	0.99	0.08	0.01
					(0.005)
Ravens Score (0–15)	8.82	3.91	7.93	4.27	0.88
					(0.725)
Employment Status					
Unemployed	0.36	0.49	0.46	0.50	-0.09
					(0.090)
Employed	0.06	0.24	0.10	0.30	-0.04
					(0.046)
Self-Employed	0.58	0.50	0.44	0.50	0.13
					(0.092)
Highest Schooling					
Secondary School	0.45	0.51	0.33	0.47	0.12
					(0.092)
Vocational Training	0.03	0.17	0.01	0.12	0.02
					(0.031)
National Diploma	0.09	0.29	0.16	0.37	-0.07
					(0.055)
University	0.42	0.50	0.49	0.50	-0.07
					(0.092)
Observations	33	33	288	288	321

Note: This table shows balancee of the baseline employee sample for the first employment round. We made 321 joboffers to of which 300 accepted the job. The table shows balance between individuals assigned to treatment arm 1 and treatment arm 2.

Table B.8: Recruitment Sample

	Mean	SD	10th Percentile	90th Percentile	Observations
Panel A: In-Person Recruitment Sample (Incentivized)					
Characteristics					
Age	27.16	7.57	18.00	38.00	821
% Female	0.42	0.49	0.00	1.00	821
Raven's Matrices	7.40	3.80	2.00	13.00	821
% Searching for a Job	0.62	0.49	0.00	1.00	821
% Attending Orientation	0.44	0.50	0.00	1.00	821
Employment Status					
Unemployed	0.26	0.44	0.00	1.00	821
Employed	0.15	0.36	0.00	1.00	821
Self-Employed	0.59	0.49	0.00	1.00	821
Highest Schooling					
Secondary School	0.63	0.48	0.00	1.00	821
Vocational Training	0.01	0.12	0.00	0.00	821
National Diploma	0.13	0.33	0.00	1.00	821
University	0.20	0.40	0.00	1.00	821
Panel B: In-Person Recruitment Sample (Unincentivized)					
Characteristics					
Age	25.65	7.12	18.00	35.00	258
% Female	0.40	0.49	0.00	1.00	258
Raven's Matrices	8.09	4.27	2.00	13.00	258
% Searching for a Job	0.89	0.32	0.00	1.00	258
% Attending Orientation	0.60	0.49	0.00	1.00	258
Employment Status					
Unemployed	0.47	0.50	0.00	1.00	258
Employed	0.11	0.32	0.00	1.00	258
Self-Employed	0.41	0.49	0.00	1.00	258
Highest Schooling					
Secondary School	0.60	0.49	0.00	1.00	258
Vocational Training	0.02	0.14	0.00	0.00	258
National Diploma	0.15	0.36	0.00	1.00	258
University	0.22	0.42	0.00	1.00	258
Panel C: Job Advertisement Sample					
Characteristics					
Age	29.53	9.38	18.00	42.00	638
% Female	0.53	0.50	0.00	1.00	638
Raven's Matrices	8.16	4.15	2.00	13.00	638
% Searching for a Job	0.99	0.09	1.00	1.00	638
% Attending Orientation	0.95	0.22	1.00	1.00	638
Employment Status					
Unemployed	0.51	0.50	0.00	1.00	638
Employed	0.10	0.30	0.00	0.00	638
Self-Employed	0.39	0.49	0.00	1.00	638
Highest Schooling					
Secondary School	0.39	0.49	0.00	1.00	638
Vocational Training	0.01	0.09	0.00	0.00	638
National Diploma	0.16	0.37	0.00	1.00	638
University	0.44	0.50	0.00	1.00	638

Note: This table shows average characteristics and statistics across our three different samples. Panel A: in-person recruited and requiring an incentive to participate in the job information session. Panel B: in-person recruited and willing to participate in the job information session without an incentive. Panel C: job advertisement sample, i.e. individuals who responded to our job advertisement.

Table B.9: Employee Sample

	Mean	SD	10th Percentile	90th Percentile	Observations
Panel A: Job Offer Acceptance					
Job offer accepted (Control Offer 1)	1.00	0.00	1.00	1.00	33
Job offer accepted (Uncertainty Offer 2)	0.95	0.22	1.00	1.00	281
Panel B: Employee Demographics					
Characteristics					
Age	30.72	9.73	20.00	44.50	300
% Female	0.54	0.50	0.00	1.00	300
% Searching for a Job	0.99	0.08	1.00	1.00	300
Employment Status					
Unemployed	0.46	0.50	0.00	1.00	300
Employed	0.08	0.27	0.00	0.00	300
Self-Employed	0.46	0.50	0.00	1.00	300
Highest Schooling					
Secondary School	0.35	0.48	0.00	1.00	300
Vocational Training	0.02	0.13	0.00	0.00	300
National Diploma	0.15	0.36	0.00	1.00	300
University	0.48	0.50	0.00	1.00	300
Panel C: Employee Work Patterns					
Total Treatment Assignments	1.03	0.98	0.00	3.00	300
Absent (0,1)	0.21	0.20	0.02	0.50	300
Hours at work	7.16	0.32	6.89	7.38	300
Time spent working	6.12	0.58	5.40	6.71	300
Images labeled per day	255.73	156.71	85.68	487.99	300
Images labeled in total	15,332.38	9,409.85	5,130.00	29,275.00	300
Time spent per image	89.96	50.73	41.88	156.16	300
Correct labels per image (0,10)	8.37	0.39	7.97	8.66	300
All labels correct (0,1)	0.10	0.04	0.04	0.14	300

Note: This table shows the acceptance rate of job offers and the characteristics of the employee sample.

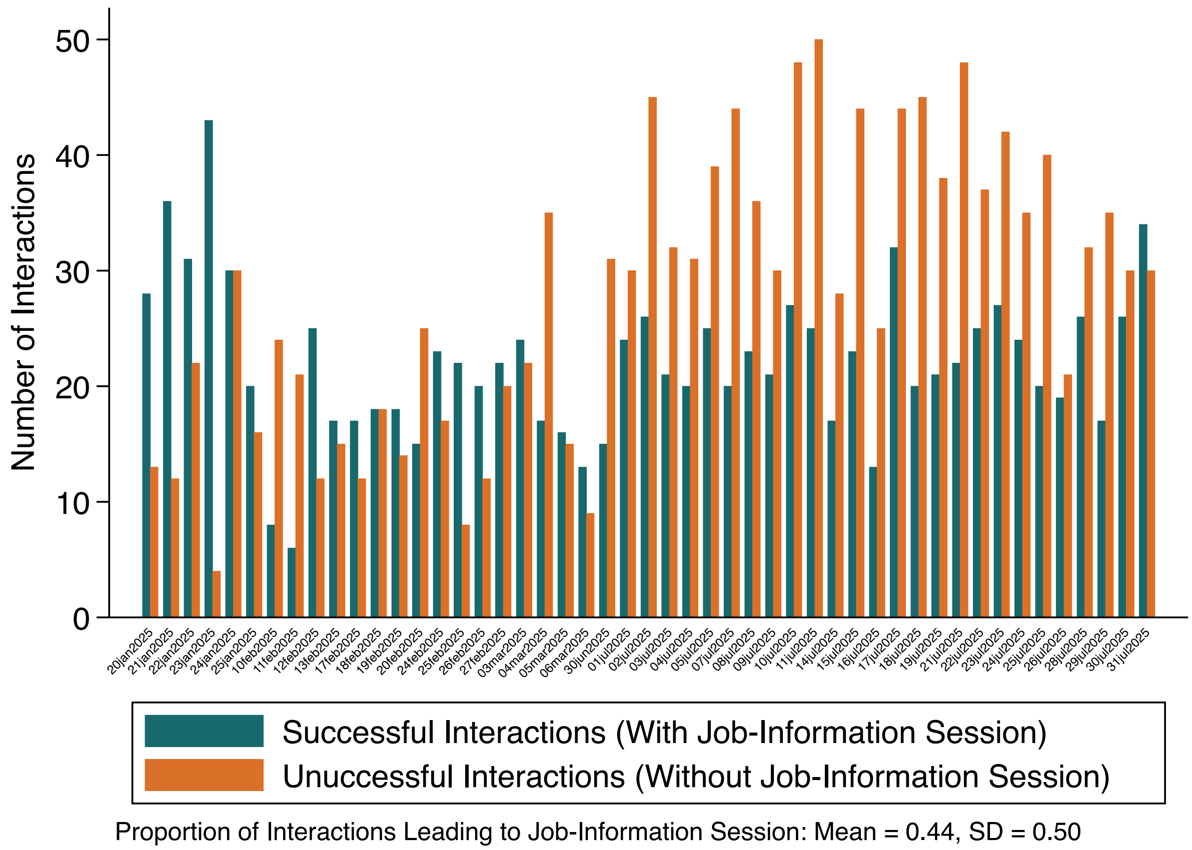
Figure B.2: Job Advertisement

NOTE: This figure shows the content of the job advertisement that we posted across Lagos.

Figure B.3: Posted Job Advertisement

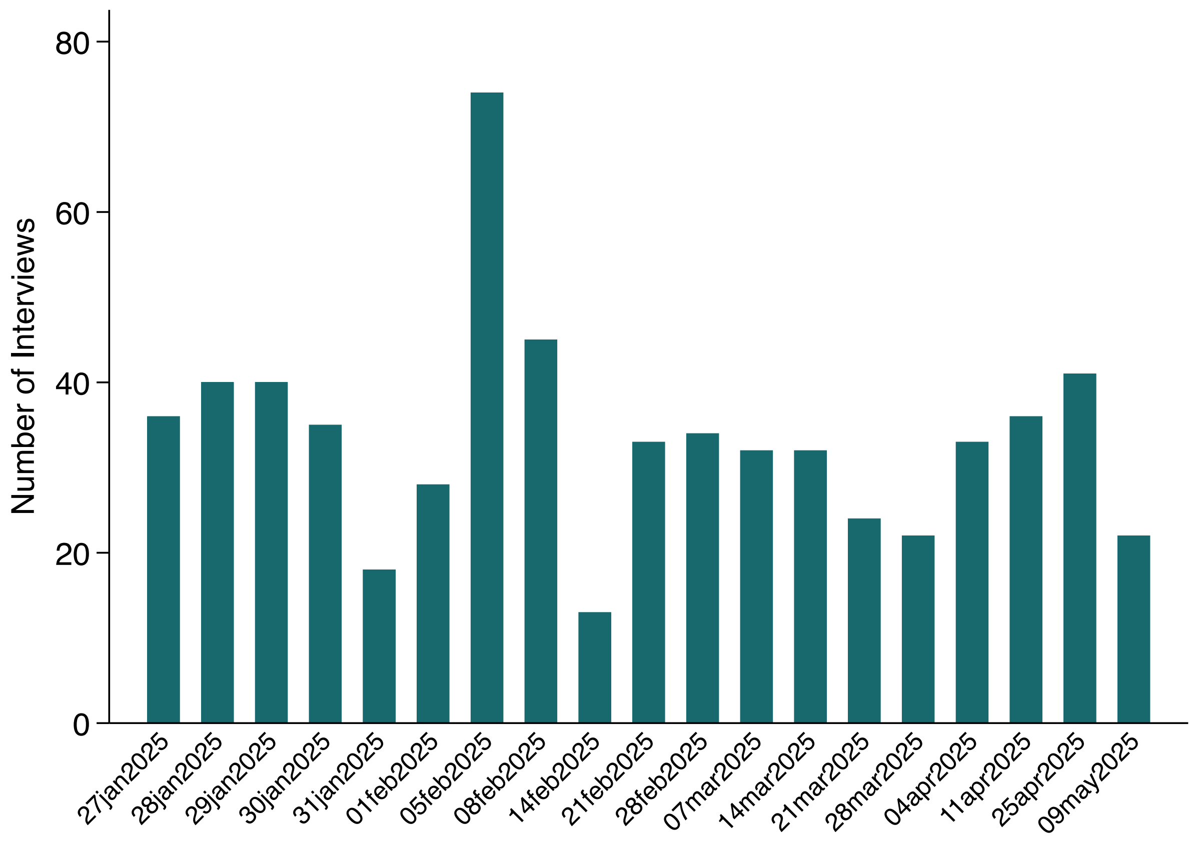
Note: This figure shows the job advertisement that we posted on a concrete wall together with other job advertisements. This figure is intended to provide an example posting of another job advertisement and demonstrate that our approach resembles common practice in this setting.

Figure B.4: In-Person Recruitment Timing



Note: This figure shows the timeline of our in-person recruitment. The teal darker teal bars show the number of interactions of our enumerators that were successful and lead to a job-information session. The lighter orange bars show the number of interactions that did not lead to a job-information session. On average, 44 percent of interactions were successful (mean = 0.44, SD = 0.50).

Figure B.5: Interview Timing



NOTE: This figure shows the timeline of job interviews with jobseekers who responded to our job advertisements. Bars indicate the number of interviews conducted per day, restricted to days on which interviews took place

C Image Labelling Task

Lego pieces are plastic building blocks made for children to play with that come in many shapes, sizes, and colors. They are designed to fit together securely, allowing the creation of all kinds of complex structures.

C.1 Creation of Images

All images used in the labeling task were generated digitally using the Lego bricks online simulator *mecabricks.com*. This simulator provides access to the complete set of Lego brick types and allows users to arrange them in a virtual 3D space. The resulting brick geometries can be exported and processed, for example, via a Python script to produce images.

We used the simulator to create 1,000 unique Lego brick geometries: 100 containing a single brick and 900 containing multiple bricks. Each geometry was imported into the image renderer *Autodesk Maya 2024*, where we generated 40 images per geometry with a plain white background, varying the camera position for each shot. This procedure generated 40,000 unique images. From this set, we randomly selected 4,000 images; we duplicated them and replaced the plain white background with a visually noisy, captcha-style background. This resulted in a final dataset of 44,000 unique images: 40,000 regular images and 4,000 'high-stakes' images for our labelling task.

Because the images were fully computer-generated, we maintained complete control over attributes such as color, spatial arrangement, and viewing angle. By generating all images digitally ourselves, we maintained complete control over every attribute—including color, spatial arrangement, and viewing angle. This allowed us to define the true classification for each image and evaluate employees' performance against it.

C.2 Labelling Categories

Employees were required to label each image according to six categories of heterogeneous complexity: (a) the color of the Lego pieces (all pieces in a given image were the same color); (b) whether any Lego pieces in the image were stacked together; (c) the total number of individual Lego pieces in the image; (d) the types of Lego pieces visible in the image (bricks, plates, bows, circles, and angles); (e) the type of Lego piece appearing most frequently in the image; and (f) whether a 2×2 Lego brick was visible in the image.

(a) The Color of the Lego Pieces. All Lego pieces in a given image were the same color. Employees were required to identify this color and select the correct option from a dropdown list of nine predefined colors: red, orange, yellow, green, teal, blue, purple, brown, and black. A response was classified as correct if the chosen color matched the color assigned to the bricks during the generation of the digital image.

(b) Whether any Lego Pieces in the Image Were Stacked Together. In most images, Lego pieces were arranged without touching each other. However, in some geometries — and consequently in some images — some or all pieces were arranged so that they appeared to be stacked together. Employees were required to determine whether any pieces in a given image were stacked and to select the correct "yes" or "no" response from a dropdown list to the question: "Are the pieces in the image stacked together?"

(c) Total Number of Individual Lego Pieces in the Image. Images contained between one and seven individual Lego pieces. Employees were required to count the number of pieces and select the correct value from a dropdown list of options ranging from one to seven.

(d) Types of Lego Pieces Visible in the Image. All pieces included in the images belonged to one of five categories: bricks, plates, bows, circles, or angles. Bricks were defined as any cubic piece. Plates were defined as any flat rectangular piece. Bows were defined as curved pieces that are not circular. Circles were defined as any circular piece. Angles were defined as any piece without curves and with at least one angle different from 90°.

An image could contain pieces from one or multiple categories. Employees were required to tick a checkbox for each type visible in the image; leaving a checkbox unticked indicated that the type was not visible. This yielded five separate responses — one for each category. For each category, we compared the employee's response with the correct classification for that image. A response was classified as correct if the checkbox selection matched the true appearance of the respective category in the given image.

(e) Type of Lego Piece Appearing Most Frequently in the Image. Employees were required to identify the type of piece that appeared most often in the image. For example, if the image contained three bricks, two plates, and one circle, the correct response would be "brick." Employees selected the answer from the list of five possible types.

If two or more types appeared the same number of times — for instance, two plates and two circles — employees were required to select the option "no single type appears most often" instead of choosing one of the five types. A response was classified as correct if the selected option from the dropdown list matched the true most-frequent type or correctly indicated that no single type appeared most often.

(f) Whether a 2×2 Lego Brick Was Visible in the Image. Employees were asked to indicate whether a specific Lego piece—a 2×2 brick—was visible in the image. This is a common Lego piece, illustrated in Appendix Figure C.1, and is relatively easy to identify. Employees selected either "yes" or "no" from a dropdown list in response to the question: "Is there a 2×2 brick in the image?"

C.3 Set up of the task.

All employees were provided with identical laptops for the task, as well as individual *Labelbox* accounts with unique login credentials that we set up in advance. Once logged in, they could begin labeling immediately. Each image was shown only once, and employees had no opportunity to return to a previously submitted image to change their answers. No feedback on performance — either absolute or relative to other employees — was provided within *Labelbox* or in any other form. Appendix Figure C.3 shows the *Labelbox* labeling interface with an example image, presented exactly as employees saw it during the task.

On the first day, employees received extensive training on the image classification task and the use of *Labelbox*. Each employee had a PDF copy of the training materials on their laptop and was encouraged to consult it whenever they had questions about a labeling task. In addition, each work location had physical examples of Lego pieces available, so employees could familiarize themselves with the different types of pieces.

C.4 Representativeness of the Lego Task

The Lego task exemplifies *data labeling*, a core activity in modern AI development. Data labeling is the process of adding descriptive information to raw data —for example, classifying objects in images or identifying the sentiment of a text. These labels provide the examples that machine-learning systems need in order to learn and make predictions about new observations. Because algorithms cannot reliably generate such labels on their own, data labeling is typically carried out by human workers around the world.

For instance, crowdsourced workers across the globe (e.g., on platforms like Appen or Scale AI) label images, text, and video for major tech firms, often in low-income countries (Chen, 2023; Larousserie, 2024; The Economist, 2025). Kässi et al. (2021) estimate that over 163 million freelancer profiles are registered on online labor platforms, with approximately 19 million having performed at least one job, and 5 million completing at least ten tasks or earning over USD 1,000. While not all of this work involves data labeling, Appen alone reports over one million data-labeling contributors globally (Chen, 2023). The growing institutionalization of this sector is also evident in the creation of organizations such as the *Data Labeling Association of Kenya*, which represents local workers engaged in annotation tasks and advocates for fair labor standards.[70]

Data labeling is among the fastest-growing forms of digital work globally, with the market for labeling solutions projected to grow at more than 20% annually through 2030 (G2 Learning Hub, 2023). The broader AI data-labeling market is forecast to reach USD 134.7 billion by 2034, up from USD 19.7 billion in 2024 (Market.us Insights, 2024).

This scale illustrates that data labeling is a ubiquitous form of digital labor. Through the Lego-image labeling task, we reproduce the core elements of large-scale labeling workflows — repetition, cognitive engagement, and precision in categorization — within a controlled and measurable setting. These features also characterize many other forms of work.

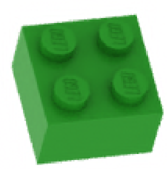
C.5 Image-Level Summary Statistics

Table C.1 summarizes the resulting dataset of 4,595,228 labelled images. Panel (a) presents data at the image-employee level, with each observation representing a single employee's labeling of an image. Employees spent an average of 68.4 seconds labeling each image and correctly assigned an average of 8.4 labels per image. Overall, this translates into an 11 percent probability that an image is labeled entirely correctly by any given employee, indicating considerable variation in labeling accuracy and speed.

Panel (b) aggregates data to the image level, with each of the 44,000 unique images representing a single observation, averaged across multiple employee labels. At this aggregated level, the average accuracy per image remains 11 percent, with each image labeled by an average of 104 different employees.

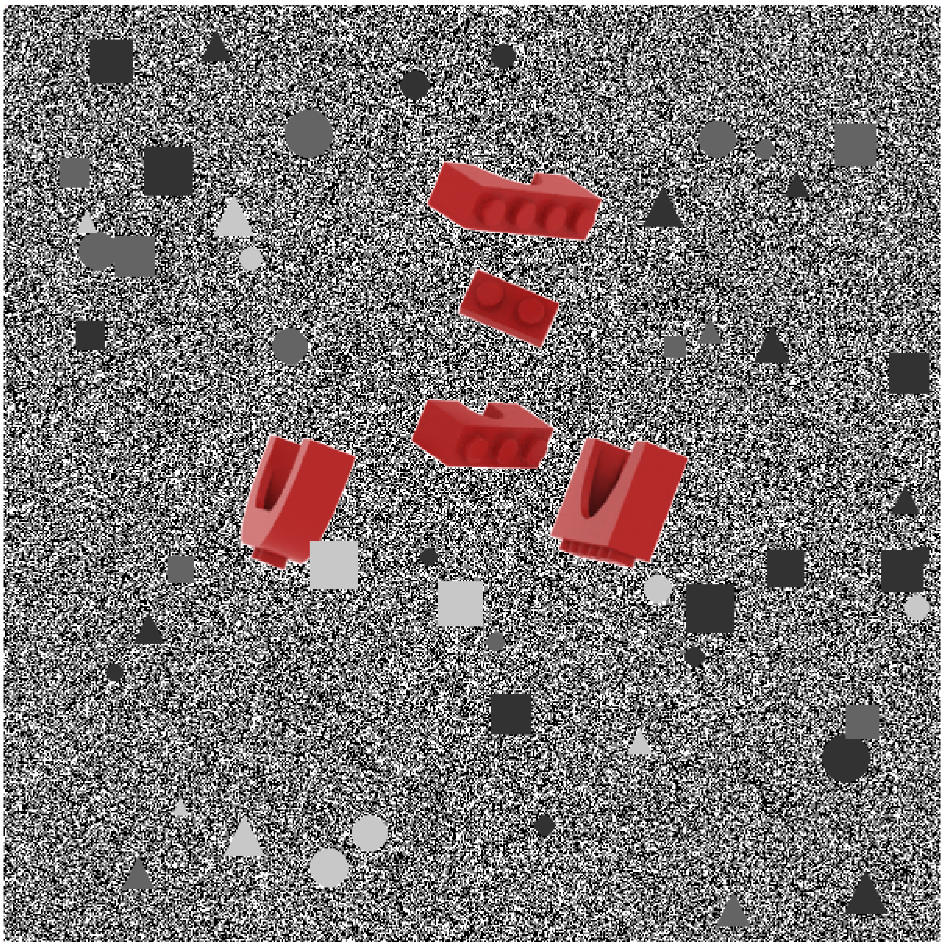
[70] See: https://datalabelers.org

Figure C.1: 2×2 **Brick**



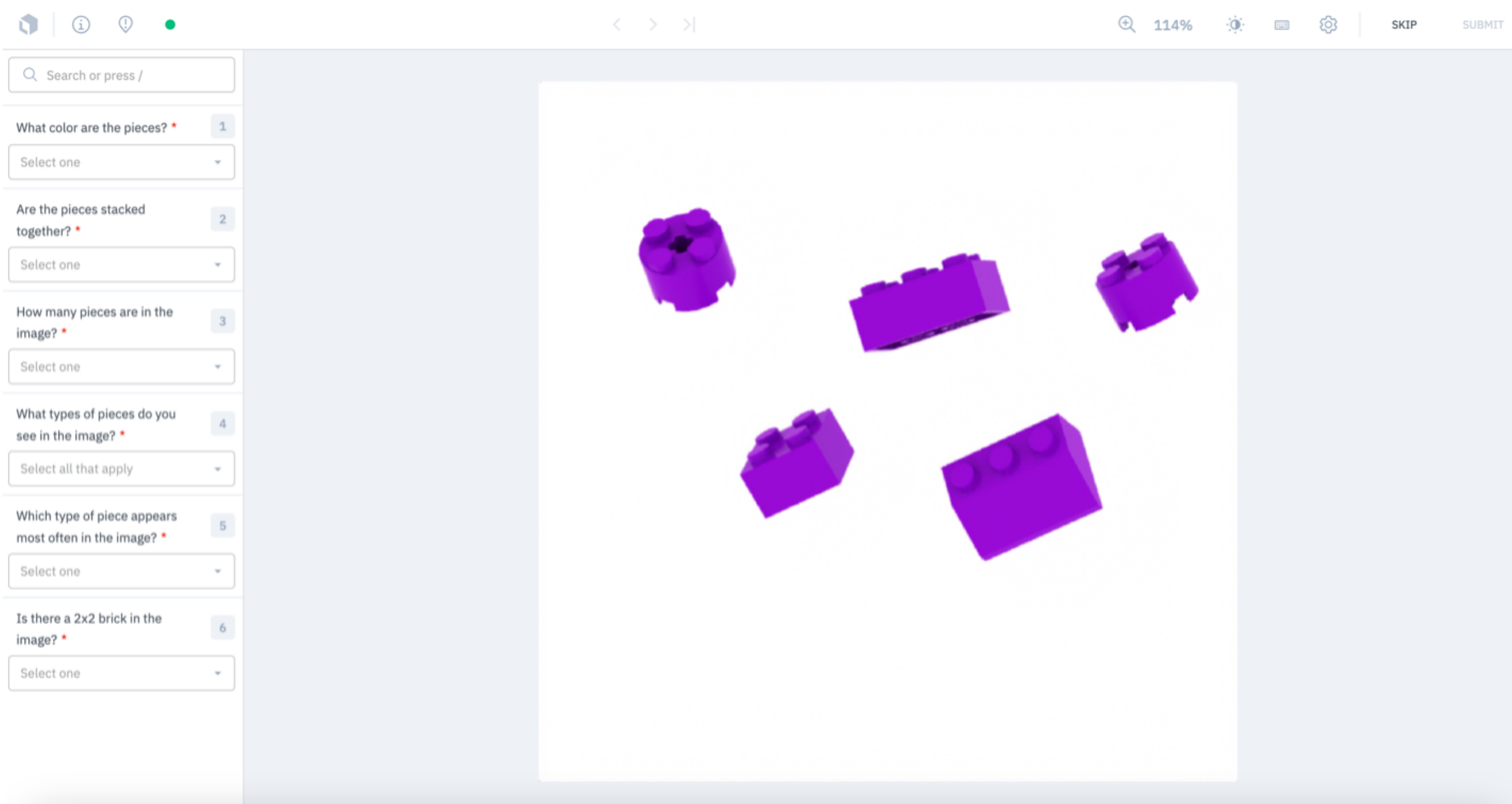
NOTE: This figure shows a 2×2 brick.

Figure C.2: Image with Captcha Style Background



Note: This figure shows an example image from the labeling task with a visually noisy, captcha-style background. Such images were designated as 'high-stakes', and employees were instructed to devote extra effort and care when labeling them.

Figure C.3: Examples Image of Lego Bricks



NOTE: This figure shows a screenshot of an example Lego bricks image with the entry boxes for the classification of the image as employees have seen during the experiment. This screenshot is taken from the data classification platform *Labelbox*, which was used during the field experiment to classify the Lego images.

Table C.1: Lego Bricks Data

	Mean	SD	10th Percentile	90th Percentile	Observations
Panel A: Image Employee Level Dataset					
Correct Labels (0–10)	8.43	1.01	7.00	10.00	4,595,228
All labels correct (0,1)	0.11	0.31	0.00	1.00	4,595,228
Time spent per image	68.44	115.78	16.00	137.00	4,595,228
Panel B: Image Level Dataset					
Correct Labels (0–10)	8.53	0.75	7.67	9.67	44,000
All labels correct (0,1)	0.11	0.27	0.00	0.74	44,000
Time spent per image	47.93	24.13	18.25	80.81	44,000
Employees	104.44	115.33	4.00	296.00	44,000

NOTE: This table summarizes the dataset of Lego images and corresponding labelling performance from the first employment round.

D Validity Checks

We perform a test for pre-trends as suggested by Borusyak et al. (2024). The idea behind their pre-trend test is to use the untreated sample only and include time dummys for pre-treatment timing. Specifically, we estimate the following regression:

$$y_{itq} = \alpha + \delta_q + \kappa_i + \mu_t + \sum_{k=g_{-1}}^{-4} \beta_k \times \text{treat}_{ik} + \varepsilon_{itq} \tag{10}$$

As before, y_{itq} refers to outcome measure y for individual i on date t and image q. We include indicators k indicating relative time periods to treatment before treatment started. The relevant time period here is a week. I.e. we include indicators for up to four weeks before treatment started. If there are no pre-trends, these coefficients should be statistically insignificant. Results of this equation are presented in the eventstudy graphs in Figure 4.

Implication for parallel trends. Assumption E1 implies a standard parallel trends assumption holds for untreated potential outcomes. Let $Y_{itq}(0)$ denote the potential outcome of individual i at time t on image q in the absence of treatment. Since Assumption E1 states that

$$g_i \perp \{Y_{itq}(g)\}_{t,g \in \mathcal{G}} \,\big|\, J_i,$$

it follows that

$$g_i \perp \{Y_{itq}(0)\}_t \,\big|\, J_i.$$

Hence, for any times s, t, any stratum j and any treatment timings $g, g' \in \mathcal{G}$,

$$\mathbb{E}[Y_{itq}(0) - Y_{isq}(0) \mid g_i = g, J_i = j] = \mathbb{E}[Y_{itq}(0) - Y_{isq}(0) \mid g_i = g', J_i = j]\,,$$

which is a standard (within-stratum) parallel trends assumption underlying the event-study estimator.

E Robustness

E.1 Alternative Definition of Productivity Measure

In this subsection, we examine labeling performance and treatment effects using a pre-registered subset of the labeling questions. The index is constructed from questions (c) and (d) described in Appendix C.2. Accordingly, the continuous index ranges from zero to six, while the binary index equals one if all six questions are answered correctly and zero otherwise. Table E.1 replicates Table 3 — our main analysis table — using the alternative productivity index. Qualitative results do not change when using the alternative native productivity measure.

Table E.1: Treatment Effects on Worker Effort for Alternative Productivity Measure

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0138	0.0013	0.0147	0.0004
	(0.0043)***	(0.0006)**	(0.0057)**	(0.0012)
	[0.0046]***	[0.0007]**	[0.0071]**	[0.0014]
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Individual	Individual	Individual	Individual
Mean of Dep. Var.	4.69	.14	4.68	.13
Q-Value	0.005	0.025	0.011	0.235

NOTE: This table reports the effect of delayed salary payments on employees' effort measures from equation (4), using the described alternative productivity indices aggregated into a single ATE. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the individual level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses. Hence, this table replicates Table 3, using our alternative productivity measures.

E.2 Spillover Concerns

A key threat to the experimental design is the possibility of spillover effects because employees with different treatment status work in the same location. In this setting, morale or other behavioral responses may spread from treated to untreated employees, potentially biasing the estimated treatment effects. We address this threat in two ways. First, we use outcomes from a location that was deliberately assigned to contain only untreated employees to impute counterfactual outcomes for treated employees. Second, within mixed-treatment locations, we explicitly designed fixed seating arrangements to create reference groups, allowing us to account for potential spillover effects within each group.

Mitigating Spillovers with an Untreated-Only Location. The first way we address potential spillover effects is by relying on untreated employees from a work location with only untreated workers. We replicate Table 3 and Figure 4, restricting the sample of untreated employees to this group. The key advantage of this approach is that these employees are certainly unaffected by treated peers, since they work in a separate location. Two disadvantages remain. First, we still rely on untreated pre-periods of employees who share a location with other treated workers, so some concerns about spillovers may remain. Second, restricting the sample of untreated employees to the ones working in the untreated-only location reduces sample size and statistical power, making the estimates noisier. Nonetheless, the results are qualitatively similar to those reported in Table 3 and Figure 4. Table E.2 presents estimates using the standard continuous and binary indices, while Table E.3 reports results with the alternative indices. Columns1-2 reproduce the estimates based on all images for reference, and Columns (3)-(4) restrict to untreated observations from the untreated-only location to impute counterfactual outcomes for the treated sample.

Table E.2: Treatment Effects: Imputations Using Untreated-Only Location

	All Images		Untreated-Only Location	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0449	0.0062	0.0330	0.0030
	(0.0164)***	(0.0027)**	(0.0138)**	(0.0019)
	[0.0176]***	[0.0029]**	[0.0185]**	[0.0029]
Observations (First Stage)	2,166,877	2,166,877	1,368,152	1,368,152
Observations (Imputed)	675,184	675,184	675,184	675,184
Individuals	297	297	217	217
SE Cluster	Individual	Individual	Individual	Individual
Mean of Dep. Var.	8.43	.11	8.44	.11
Q-Value	0.010	0.016	0.026	0.047

Note: This table reports the effect of delayed salary payments on employees' effort measures estimated using equation (4), aggregated into a single ATE. Columns (1)-(2) replicate the results from Table 3, while columns (3)-(4) use untreated employees from the untreated-only location to impute counterfactual outcomes for treated employees. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the individual level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses.

Table E.3: Treatment Effects: Imputations Using Untreated-Only Location (Alternative Productivity Measure)

	All Images		Untreated-Only Location	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0138	0.0013	0.0169	0.0016
	(0.0043)***	(0.0006)**	(0.0043)***	(0.0007)**
	[0.0046]***	[0.0007]**	[0.0057]***	[0.0009]**
Observations (First Stage)	2,166,877	2,166,877	1,368,152	1,368,152
Observations (Imputed)	675,184	675,184	675,184	675,184
Individuals	297	297	217	217
SE Cluster	Individual	Individual	Individual	Individual
Mean of Dep. Var.	4.69	.14	4.69	.14
Q-Value	0.005	0.025	0.001	0.035

NOTE: This table reports the effect of delayed salary payments on employees' effort measures estimated from equation (4), using our alternative productivty indices and aggregated into a single ATE. Columns (1)-(2) replicate the results from Table 3, while columns (3)-(4) use untreated employees from the untreated-only location to impute counterfactual outcomes for treated employees. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the individual level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses.

Mitigating Spillovers with Designed Reference Groups. Our second approach to addressing potential spillover effects is to create a clearly defined reference group for each employee, which allows us to control for and estimate spillovers within a work location. To construct these reference groups credibly within an open-space work location, we relied on three design features: fixed seating arrangements, restrictions on communication across groups, and staggered work schedules. First, employees were seated in groups of four around a single table, with fixed seats that remained unchanged throughout the employment period, ensuring that each individual interacted with the same peers for the duration of the job. Second, we restricted communication across tables by prohibiting conversation during work, generally allowing only one person at a time to use the washroom, and providing bottled water at each table to eliminate the need for a shared water fountain. Third, we implemented staggered start, end, and break times across tables to prevent crowding and mingling when employees arrived or left work and during break periods. We provide evidence that these measures successfully created clearly defined reference groups in Figure E.1.

Panel (a) shows that there was virtually no communication across tables, based on our daily supervisory staff survey (see Appendix I for survey details). Panel (b) demonstrates that employees were well acquainted with their table peers but not with workers at other tables, as measured in our Peer Recognition survey (see Appendix I for survey details). Together, these patterns suggest that the peer groups functioned as intended and that spillovers across groups were minimal.

We then use the reference groups to explicitly control for potential spillover effects. We continue to rely on the imputation estimator of Borusyak et al. (2024), which requires one key assumption about the nature of spillovers. Because the estimator uses untreated observations to estimate fixed effects, we must assume that spillover effects from treated employees affect untreated and treated co-workers in the same way. In this setting, that assumption is reasonable: when working alongside a peer who experiences unpaid salaries, it is plausible that the effect on others does not depend on whether those others are themselves treated.

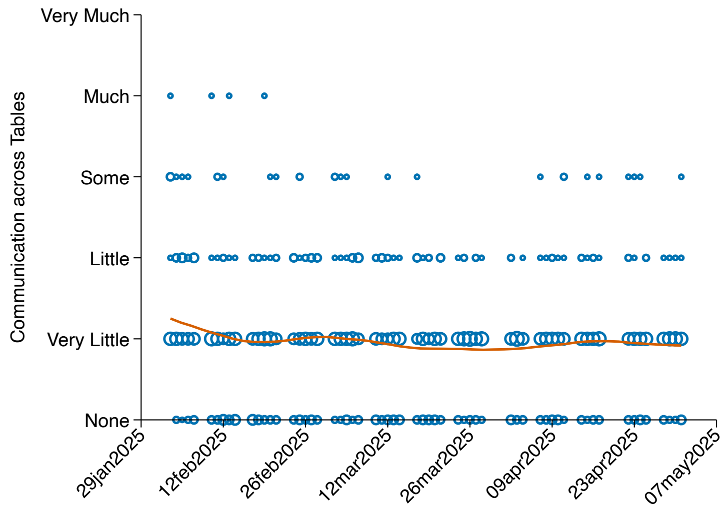
Following the notation introduced in Section 6.1, we adapt the approach of Vazquez-Bare (2023) to set up the estimation of spillover effects. Specifically, we estimate the following modified version of equation (3):

$$y_{itq} = \alpha + \delta_q + \kappa_i + \mu_t + \lambda \mathbb{1}(S_{igt} > 0) + \varepsilon_{itq} \tag{11}$$

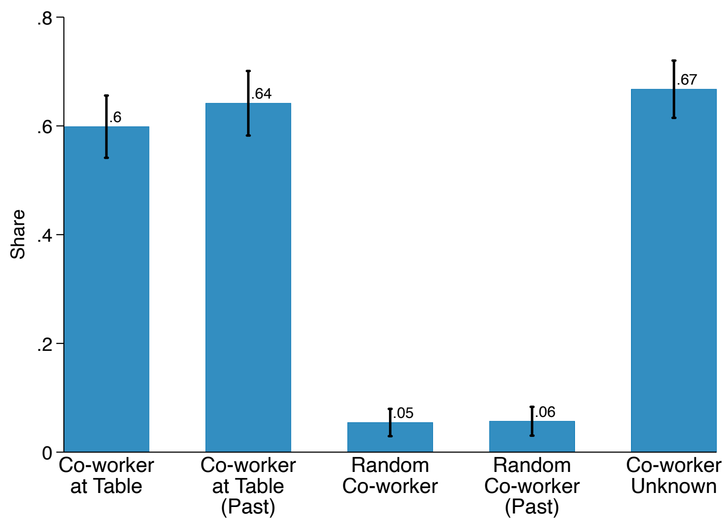
via OLS on $(i, t) \in \mathcal{S}_0$ (i.e., untreated observations only). The parameters δ_q, κ_i, and μ_t denote vectors of fixed effects as before. We additionally include S_{igt}, an indicator equal to one if individual i in reference group g is exposed to treated individuals at time t. This can include any number of treated co-workers (see Vazquez-Bare (2023) for further discussion). As before, we then use the estimated parameters from equation 11 to construct counterfactual outcomes for treated individuals. The only difference is that our model for y_{itq} now explicitly accounts for spillover effects. This adjustment allows us to distinguish the direct impact of salary delays on treated workers from indirect effects transmitted through their peers, under the assumption that spillover effects are homogeneous across untreated and treated employees.

Figure E.1: Reference Group Validation

(a) Communication Across Tables



(b) Acquaintance With Co-Workers



NOTE: This figure provides suggestive evidence validating our construction of the reference groups. Panel (a) displays Likert scale responses from the daily supervisory staff survey on communication across tables during a given workday, ranging from "None" (1) to "Very Much" (6). The panel also plots a kernel density estimate using an Epanechnikov kernel with a bandwidth of five days, based on the numeric values of the Likert scale. Panel (b) reports employees' acquaintance with co-workers. The first two bars show employees' knowledge about the treatment status of their table peers (whether they were currently owed salary or had been owed in the past), with a large share indicating familiarity. In contrast, just a very small share reported knowing the treatment status of a randomly selected co-worker seated at another table (bars three and four), and the vast majority even reported not knowing this co-worker at all (bar five). We also display 95 percent confidence intervals around the sample means.

Table E.4: **Treatment Effects: Clustering at the Table-Level**

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.045	0.006	0.048	0.008
	(0.017)**	(0.003)**	(0.016)***	(0.003)***
	[0.020]**	[0.003]**	[0.019]***	[0.004]***
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Table	Table	Table	Table
Mean of Dep. Var.	8.43	.11	8.41	.11
Q-Value	0.016	0.022	0.011	0.011
Spillover Control	No	No	No	No

Note: This table reports the effect of delayed salary payments on employees' effort measures estimated using equation (4), aggregated into a single ATE. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the table level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses. Hence, this table replicates Table 3, clustering standard errors at the table-level.

Table E.5: Treatment Effects: Clustering at the Table-Level and Accounting for Spillover Effects

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.044	0.006	0.047	0.007
	(0.017)**	(0.003)**	(0.016)***	(0.003)***
	[0.019]**	[0.003]**	[0.021]***	[0.004]***
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Table	Table	Table	Table
Mean of Dep. Var.	8.43	.11	8.41	.11
Q-Value	0.017	0.022	0.011	0.011
Spillover Control	Yes	Yes	Yes	Yes

NOTE: This table reports the effect of delayed salary payments on employees' effort measures where we account for potential spillover effects as described in equation (11), aggregated into a single ATE. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the table level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses. Hence, this table replicates Table 3, clustering standard errors at the table-level and accounting for spillover effects.

Table E.6: Treatment Effects: Clustering at the Table-Level (Alternative Productivity Measure)

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0138	0.0013	0.0147	0.0004
	(0.004)***	(0.001)**	(0.006)***	(0.001)
	[0.005]***	[0.001]**	[0.007]***	[0.001]
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Table	Table	Table	Table
Mean of Dep. Var.	4.69	.14	4.68	.13
Q-Value	0.003	0.022	0.011	0.222
Spillover Control	No	No	No	No

Note: This table reports the effect of delayed salary payments on employees' effort measures estimated using equation (4), using our alternative productivity measure and aggregated into a single ATE. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the table level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses. Hence, this table replicates Table E.1, clustering standard errors at the table-level.

Table E.7: Treatment Effects: Clustering at the Table-Level and Accounting for Spillover Effects (Alternative Productivity Measure)

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0139	0.0013	0.0148	0.0005
	(0.004)***	(0.001)**	(0.006)***	(0.001)
	[0.005]***	[0.001]**	[0.007]***	[0.001]
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Table	Table	Table	Table
Mean of Dep. Var.	4.69	.14	4.68	.13
Q-Value	0.003	0.022	0.011	0.197
Spillover Control	Yes	Yes	Yes	Yes

Note: This table reports the effect of delayed salary payments on employees' effort measures where we account for potential spillover effects as described in equation (11), using our alternative productivty measure and aggregated into a single ATE. Treatment effects are estimated at the worker-image level using the Borusyak et al. (2024) imputation estimator. Daily treatment effects are aggregated into a single parameter. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the table level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. We report sharpened q-values accounting for multiple hypothesis testing across our four productivity measures, based on the p-values in parentheses. Hence, this table replicates Table E.1, clustering standard errors at the table-level and accounting for spillover effects.

E.3 Alternative Weighting

In this subsection, we examine labeling performance and treatment effects using an alternative weighting scheme when aggregating image-employee-level treatment effects. Specifically, we weight each image inversely proportional to the number of images labeled by that employee on a given workday. This approach gives equal weight to each employee on that workday, rather than to each individual labeled image.

Table E.8: Treatment Effects on Worker Effort for Alternative Weighting

	All Images		High-Stakes Images	
	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0389	0.0055	0.0347	0.0071
	(0.0198)**	(0.0026)**	(0.0210)*	(0.0032)**
	[0.0176]**	[0.0023]**	[0.0189]*	[0.0031]**
Observations (First Stage)	2,166,877	2,166,877	196,994	196,994
Observations (Imputed)	675,184	675,184	61,209	61,209
Individuals	297	297	297	297
SE Cluster	Individual	Individual	Individual	Individual
Mean of Dep. Var.	8.43	.11	8.41	.11
Q-Value	0.071	0.071	0.119	0.119

NOTE: This table reports the effect of delayed salary payments on employees' effort measures estimated using equation (4), aggregated into a single ATE using alternative weights that are inversely proportional to the number of images labelled by a given employee on a given day. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the individual level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. q-values accounting for multiple hypothesis testing across our four productivity measures are reported, based on the p-values in parentheses. Hence, this table replicates Table 3 with different weights.

E.4 Treatment Saturation

This section shows that the results are not meaningfully affected by the level of treatment saturation within a work location. In most locations, both treated and untreated employees were present. While we argue that spillovers are unlikely to be a concern (see Appendix E.2), one might still be concerned that treatment effects differ when all employees in a location are treated versus when only a subset are. We address this concern by leveraging the first instance of treatment, when one work location happened to include only treated employees. We compare outcomes for this group during that pay cycle to those of employees in locations where only a subset were treated. Although the estimates are imprecise due to limited statistical power and the use of a single pay cycle, the results indicate qualitatively similar treatment effects, with the same sign across both groups. We interpret this as evidence that our main results are unlikely to be overly sensitive to the share of treated employees within a location.

Table E.9: Treatment Saturation

	Continuous Index		Binary Index	
	Full Treatment Saturation	Partial Treatment Saturation	Full Treatment Saturation	Partial Treatment Saturation
ATE (Pay Cycle One)	0.015	0.019	0.006	0.001
	(0.020)	(0.016)	(0.004)	(0.003)
Observations	1,607,926	1,665,507	1,607,926	1,665,507
SE Cluster	Individual	Individual	Individual	Individual
P-Value	0.830		0.210	

NOTE: This table compares treatment effects from the first pay cycle between employees in the location where all workers were treated and those in locations with partial treatment. While the coefficients cannot be statistically distinguished, the estimates are imprecise because the sample is split and based on a single pay cycle. The treatment effects are, however, qualitatively similar.

E.5 Treatment Effects by Image Difficulty

In this subsection, we examine whether treatment effects are larger in magnitude when the sample is restricted to the most difficult images. To identify difficult images, we consider only untreated observations and apply our usual restriction of discarding the first 500 images for each employee, given the steep initial learning curve. We then calculate the average number of correct labels for each image. Because this sub-analysis was not pre-registered and multiple cutoffs are plausible, we report estimation results for the most difficult 10 percent, 25 percent, and 50 percent of images. We define the most difficult images based on labeling performance among untreated individuals only. For this subset, we compute the average performance for each image and rank images accordingly. The most difficult 10 percent are defined as the lowest-ranked 10 percent of images, and we proceed analogously for alternative cutoffs.

Table E.10 presents the results, showing that treatment effects on the continuous index are larger for the more difficult images. The coefficient of 0.0539 in column 1, estimated on the most difficult 10 percent, indicates that treated employees achieved 0.0539 additional correct labels per image. Relative to the untreated baseline of 7.06 correct labels, this corresponds to an improvement of 0.76 percent. Coefficients in columns 3 and 5 can be interpreted in the same way and correspond to treatment effects of 0.73 and 0.57 percent, respectively. For reference, the ATE using all images was 0.53 percent.

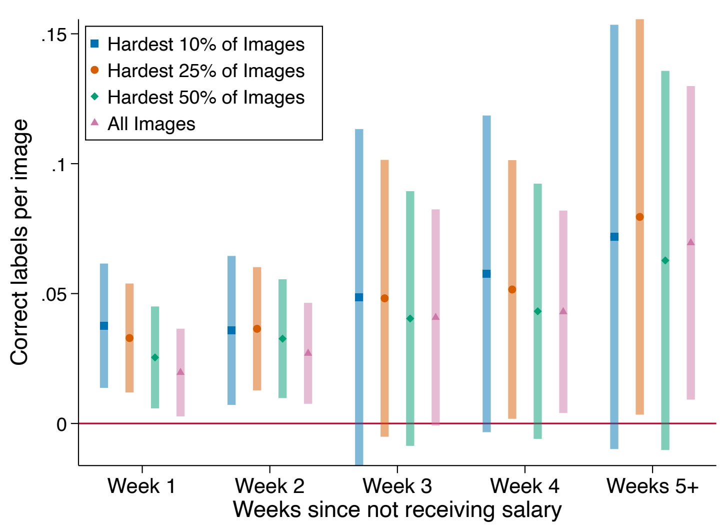
Coefficients on the binary index are all positive but statistically insignificant. However, this reflects the very low share of flawlessly labeled images: only one percent in the most difficult 10 and 25 percent, and just two percent in the most difficult 50 percent. With almost everyone making at least one mistake, it is unsurprising that treatment effects on the binary index cannot be detected statistically.

Table E.10: Treatment Effects on Worker Effort for Difficult Images

	Most Difficult 10 Percent		Most Difficult 25 Percent		Most Difficult 50 Percent	
	Continuous Index	Binary Index	Continuous Index	Binary Index	Continuous Index	Binary Index
ATE	0.0539	0.0002	0.0547	0.0001	0.0445	0.0001
	(0.0230)**	(0.0003)	(0.0205)***	(0.0001)	(0.0197)**	(0.0002)
	[0.0272]**	[0.0004]	[0.0271]***	[0.0002]	[0.0305]**	[0.0002]
Observations (First Stage)	199,276	199,276	496,231	496,231	975,265	975,265
Observations (Imputed)	42,053	42,053	114,484	114,484	230,584	230,584
Individuals	297	297	297	297	297	297
SE Cluster	Individual	Individual	Individual	Individual	Individual	Individual
Mean of Dep. Var.	7.06	.001	7.451	.001	7.836	.002
Q-Value	0.084	0.479	0.012	0.224	0.037	0.407

Note: This table reports the effect of delayed salary payments on employees' effort measures estimated using equation (4) for particularly hard images, aggregated into a single ATE. We determine the most difficult images ex-post by considering the labelling performance of the untreated group and ranking the images by average labelling performance. Standard errors based on the conservative variance estimator of Borusyak et al. (2024), clustered at the individual level and aggregated using the delta method, are shown in parentheses. Bootstrap standard errors with 500 replications, following Liu et al. (2024), are reported in brackets. 'First-stage observations' refer to the number of observations used to estimate equation (3), while 'imputed observations' indicate the number of counterfactual outcomes generated to estimate treatment effects. q-values accounting for multiple hypothesis testing across our four productivity measures are reported, based on the p-values in parentheses.

Figure E.2: Treatment Effects on Worker Effort for Difficult Images

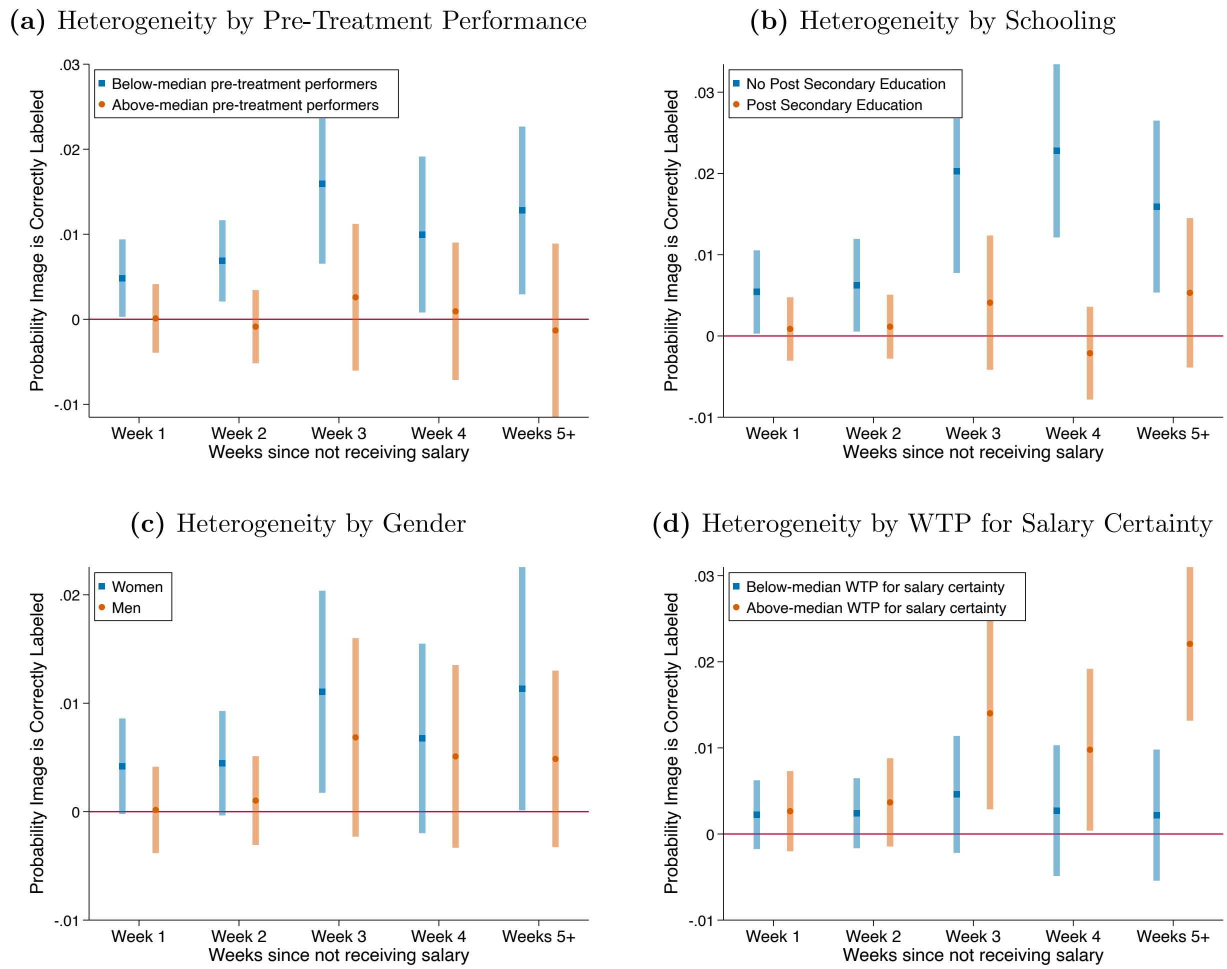


Note: This Figure visualizes the continuous effort measure estimates from Table E.10 and additionally illustrates the effort response estimates using all images.

E.6 Heterogeneity of Treatment Effects

In this section, we provide complementary results to Figure 5. We first show that the heterogeneity pattern is robust when using the binary index rather than the continuous index. We then report overall averages and formally test the null hypothesis that the groups used to split the sample perform equally.

Figure E.3: Treatment Effect Heterogeneity (Binary Index)



(a) Heterogeneity by Pre-Treatment Performance

(b) Heterogeneity by Schooling

(c) Heterogeneity by Gender

(d) Heterogeneity by WTP for Salary Certainty

Note: This figure illustrates treatment effect heterogeneity using the binary index. Panel (a) presents heterogeneity by pre-treatment performance, splitting the sample into above- and below-median performers. Panel (b) shows heterogeneity by post-secondary schooling status. Panel (c) reports heterogeneity by gender, and Panel (d) illustrates heterogeneity by WTP for higher salary certainty.

Table E.11: Average Treatment Effects of Salary Delays by Outside Option

	Performance		Schooling		Gender		WTP	
	Continuous Index	Binary Index	Continuous Index	Binary Index	Continuous Index	Binary Index	Continuous Index	Binary Index
Low Outside Option	0.078***	0.011***	0.090***	0.014***	0.053**	0.008**		
	(0.023)	(0.003)	(0.027)	(0.004)	(0.021)	(0.003)		
High Outside Option	-0.005	0.000	0.027	0.002	0.038**	0.004		
	(0.011)	(0.003)	(0.017)	(0.003)	(0.018)	(0.003)		
Low WTP							0.036*	0.003
							(0.021)	(0.003)
High WTP							0.057***	0.012***
							(0.017)	(0.003)
Observations (First Stage)	2,166,877	2,166,877	2,166,877	2,166,877	2,166,877	2,166,877	2,166,877	2,166,877
Observations (Imputed)	675,184	675,184	675,184	675,184	675,184	675,184	675,184	675,184
Individuals	297	297	297	297	297	297	297	297
SE Cluster	Individual	Individual	Individual	Individual	Individual	Individual	Individual	Individual
Mean of Dep. Var.	8.43	.11	8.43	.11	8.43	.11	8.43	.11
p-value (Low=High)	<0.001	.013	.034	.004	.581	.291	.397	.01

NOTE: This table complements Figure 5. We aggregate treatment effects into overall averages and test the hypothesis of equal performance between the groups.

F Theory

F.1 Assumptions and Timing

This section lays out all formal assumptions underlying the model and derives the first-order condition for the worker's problem.

Assumption M1 (Firm types and prior)**.** *Each firm is permanently of type $d \in \{S, N\}$. Type-S firm* never *repays wages; type-N firm pays wages and wage arrears unless facing liquidity shocks. Newly matched worker assigns prior probability $\lambda_0 \in (0, 1)$ that their employer is type-S.*

Assumption M2 (Geometric liquidity shocks)**.** *Each period, a type-N firm may independently experience a liquidity shock lasting a random number of periods $L \sim \text{Geom}(\rho)$, with exit probability $\rho \in (0, 1)$. During a shock the firm withholds the contractual wage $w > 0$. Shocks are i.i.d. across firms.*

Assumption M3 (Arrears Recovery)**.** *If the match terminates while arrears B are owed, the worker recovers each outstanding dollar with probability $\phi \in (0, 1)$. Hence, the expected transfer upon separation is ϕB.*

Assumption M4 (Worker Preferences)**.** *Denote any potential transfer from the firm as c. Per-period utility is $u(c) - \psi(e)$ with $u' > 0$, $u'' \leq 0$, $\quad \psi' > 0$, $\psi'' > 0$.*

Assumption M5 (Match-Survival)**.** *Conditional on effort e the job survives to the next period with probability $p(e)$ where $p' > 0$, $p'' \leq 0$, and $p(0) = 0$.*

Timing and Information. At the start of pay cycle t, the worker knows the arrears due this period B_t and holds a belief λ_t about the firm being strategic. First, the payment/liquidity state realizes: a type-S (strategic) firm never pays, while a type-N firm pays in full unless in a liquidity shock. The payment decision $\chi_t \in \{0, 1\}$ is then observed and arrears evolve to next period according to $B_{t+1} = (1 - \chi_t)(B_t + w)$. Upon observing no payment ($\chi_t = 0$), the worker updates the belief to λ_{t+1} (there is no updating about the liquidity parameter ρ, which is common knowledge). Given the updated state (B_{t+1}, λ_{t+1}), the worker chooses effort e_t, which determines the probability $p(e_t)$ that the match survives into period $t+1$; with probability $1-p(e_t)$ the match breaks and outstanding arrears are recovered with probability ϕ. Thus, χ_t governs the payment event (and belief update), $p(e_t)$ governs continuation, ϕ governs expected recovery at separation, and ρ governs the liquidity-shock process but is not learned about over time.

F.2 Derivations

Derivation of First-Order-Condition. Let the period-t value (given (B_t, λ_t)) be

$$V_t \;=\; \max_{e_t}\Big\{u\big(\chi_t(B_t+w)\big) - \psi(e_t) + \beta\big[\,p(e_t)\,\mathbb{E}V_{t+1} + (1-p(e_t))\big(V^{\text{out}} + \phi\,u(B_{t+1})\big)\,\big]\Big\}.$$

Given the timing (payment χ_t realizes before effort), $B_{t+1} = (1-\chi_t)(B_t+w)$ is independent of e_t. Thus, the only $\psi(e_t)$ and $p(e_t)$ depend on e_t, so

$$\begin{aligned}
\frac{\partial V_t}{\partial e_t} &= \frac{\partial}{\partial e_t}\Big\{u\big(\chi_t(B_t+w)\big) - \psi(e_t) + \beta\big[\,p(e_t)\,\mathbb{E}V_{t+1} + (1-p(e_t))\big(V^{\text{out}} + \phi\,u(B_{t+1})\big)\,\big]\Big\} \\
&= -\psi'(e_t) \;+\; \beta\,p'(e_t)\Big[\,\mathbb{E}V_{t+1} \;-\; \big(V^{\text{out}} + \phi\,u(B_{t+1})\big)\Big].
\end{aligned}$$

Optimality requires that the derivative of the value function with respect to effort equals zero, $\frac{\partial V_t}{\partial e_t} = 0$, which implies

$$\psi'(e_t^*) \;=\; \beta\,p'(e_t^*)\Big[\,\mathbb{E}V_{t+1} \;-\; \big(V^{\text{out}} + \phi\,u(B_{t+1})\big)\Big]. \tag{12}$$

It follows immediately that if $\mathbb{E}V_{t+1} \;<\; \big(V^{\text{out}} + \phi\,u(B_{t+1})\big)$ then $e_t^* = 0$. That is, if the worker's outside option is sufficiently high, or if enforcement mechanisms are sufficiently strong, the worker's optimal response is to exert zero effort as soon as a wage payment is not honored by the employer.

Optimal Effort with Growing Arrears. Let

$$\Phi(e_t, B_t) := \psi'(e_t) \; - \; \beta\, p'(e_t)\Big[\mathbb{E}V_{t+1} \; - \; \big(V^{\text{out}} + \phi\, u(B_{t+1})\big)\Big]$$

and as before the law of motion for arrears is

$$B_{t+1} = (1 - \chi_t)(B_t + w).$$

At an interior optimum $e_t^*(B_t)$, the FOC is $\Phi(e_t^*(B_t), B_t) = 0$. Implicitly differentiating with respect to B_t gives

$$\frac{de_t^*}{dB_t} = -\,\frac{\partial\Phi/\partial B_t}{\partial\Phi/\partial e_t}. \tag{13}$$

We compute the two partial derivatives in (13) in turn.

Step 1: $\partial\Phi/\partial B_t$. Using $B_{t+1} = (1 - \chi_t)(B_t + w)$ and the chain rule,

$$\begin{aligned}
\frac{\partial\Phi}{\partial B_t} &= -\,\beta\, p'(e_t)\, \frac{\partial}{\partial B_t}\Big\{\mathbb{E}V_{t+1} - \big(V^{\text{out}} + \phi\, u(B_{t+1})\big)\Big\} \\
&= -\,\beta\, p'(e_t)\,\Big(\mathbb{E}\big[V_B(B_{t+1})\big] \; - \; \phi\, u'(B_{t+1})\Big)\,\frac{\partial B_{t+1}}{\partial B_t} \\
&= -\,\beta\, p'(e_t)\,\Big(\mathbb{E}\big[V_B(B_{t+1})\big] \; - \; \phi\, u'(B_{t+1})\Big)\,(1 - \chi_t),
\end{aligned}$$

where $V_B(\cdot) \equiv \partial V(\cdot)/\partial B$ is the derivative of the worker's value with respect to arrears, i.e., the marginal continuation value as arrears increase.

Step 2: $\partial\Phi/\partial e_t$. From the definition of Φ,

$$\frac{\partial\Phi}{\partial e_t} = \psi''(e_t) \; - \; \beta\, p''(e_t)\Big[\mathbb{E}V_{t+1} - \big(V^{\text{out}} + \phi\, u(B_{t+1})\big)\Big].$$

Combining Steps 1-2. Evaluating (13) at the optimum $e_t^* = e_t^*(B_t)$ yields

$$\frac{de_t^*}{dB_t} = \frac{\beta\, p'(e_t^*)\Big(\mathbb{E}\big[V_B(B_{t+1})\big] \; - \; \phi\, u'(B_{t+1})\Big)(1-\chi_t)}{\psi''(e_t^*) - \beta\, p''(e_t^*)\Big[\mathbb{E}V_{t+1} - \big(V^{\text{out}} + \phi\, u(B_{t+1})\big)\Big]}. \tag{14}$$

By Assumption M4 ($\psi'' > 0$) and Assumption M5 ($p'' \leq 0$), the denominator in (14) is positive at an interior solution. Hence, the sign of $\frac{de_t^*}{dB_t}$ is governed by the numerator. Here, we have $p' > 0$ by Assumption M5 and β is also positive. Moreover, only realizations with $\chi_t = 0$ (nonpayment) make B_{t+1} depend on B_t. This also aligns with the problem at hand, since we study the situation of sustained non-payment. Evaluating (14) at $\chi_t = 0$ and $B_t = 0$ (so $B_{t+1} = w$) therefore yields

$$\left.\frac{de_t^*}{dB_t}\right|_{B_t=0,\ \chi_t=0} \gtrless 0 \quad \Longleftrightarrow \quad \mathbb{E}[V_B(w)] \ \gtrless \ \phi\, u'(w). \tag{15}$$

Equation (15) states that effort rises (falls) with additional arrears near $B = 0$ if and only if the expected marginal continuation value of an extra unit of arrears exceeds (falls short of) the marginal utility from expected recovery upon separation. This formalizes the ambiguity of the initial effort response.

F.3 Estimation of Belief Parameters

We estimate the belief parameters λ_0 and ρ that govern workers' perceived likelihood of being matched with a non-paying firm and the rate at which such beliefs update when wages remain unpaid. Recall from the theoretical framework that, conditional on observing nonpayment, beliefs evolve according to

$$\lambda_t(\lambda_0, \rho) = 1 - \frac{(1-\lambda_0)\exp(-\rho B_t/w)}{\lambda_0 + (1-\lambda_0)\exp(-\rho B_t/w)}, \tag{16}$$

where B_t denotes accumulated arrears and w the contractual wage. We obtain parameter estimates by minimizing the nonlinear least squares objective

$$\min_{\lambda_0,\rho} \sum_i \left[\lambda_i^{\text{obs}} - \lambda_t(\lambda_0, \rho)\right]^2, \tag{17}$$

where λ_i^{obs} denotes the observed belief from survey responses. The resulting estimates $\hat{\lambda}_0$ and $\hat{\rho}$ describe the prior probability of facing a strategic (non-paying) firm and the sensitivity of belief updating with respect to the duration of nonpayment, respectively. Appendix Table F.1 reports the estimation results.

Table F.1: Workers' Belief Parameters

	(1)
λ_0	0.41
	(0.01)***
ρ	0.14
	(0.01)***
Observations	3,837

NOTE: This table presents estimates of equation (17) based on our survey data. We ask respondents about their beliefs regarding the likelihood of facing a firm that will not pay as agreed, as well as their expectations under hypothetical scenarios of missed payments and the perceived probability of being paid in the future.

F.4 Heterogeneity Assumptions

Assumption M6 (Outside-Option Heterogeneity)**.** *Worker type θ is drawn from cumulative distribution Θ on $[0, 1]$. The value of the outside option satisfies $V^{\text{out}}(\theta) = g(\theta)$ with $g'(\theta) > 0$.*

Assumption M7 (Outside Option as Permanent Income Stream)**.** *The worker's outside option provides a permanent income stream that yields a geometric progression of per-period utility flows based on ability θ. Specifically, upon separation the worker receives deterministic* *income $g(\theta)$ each period, generating a present value*

$$V^{\text{out}}(\theta) = \sum_{t=0}^{\infty} \beta^t g(\theta) = \frac{g(\theta)}{1-\beta},$$

where $g'(\theta) > 0$ and $\beta \in (0,1)$ is the worker's discount factor.

Assumption M8 (Risk Preference Heterogeneity)**.** *Workers differ in their degree of absolute risk aversion. Each worker is characterized by a risk parameter $a \in (0,1)$. Preferences are exponential with constant absolute risk aversion (CARA):*

$$u_a(c) = -\exp(-ac),$$

so that the coefficient of absolute risk aversion equals a. Higher values of a correspond to greater risk aversion, and a is independent of the outside-option type θ and other worker characteristics.

F.5 Endogenizing the Firm's Decision

We extend the simple framework with a version that endogenizes the firm's decision.

Environment (Formal Description). Time t is discrete and denotes pay cycles. The economy is populated with risk-averse workers and risk-neutral firms, each employing at most one worker. Firms maximize the present value of profits and are permanently characterized by a type $\theta = (A, r)$ drawn from F_Θ, where A denotes productivity and r the interest rate (cost of borrowing). Workers are of type ω and have outside option V^{out}. Firms offer workers a common contractual wage $w > 0$, but each pay cycle decide what fraction $\chi_t \in [0,1]$ of the owed amount to actually pay. At the start of pay cycle t the firm-worker pair is in state $(B_t, D_t, C_t) \in \mathbb{R}^3_+$, where B_t denote wage arrears owed to a worker, D_t a firm's debt stock, and C_t cash on hand (cash carries zero return). Nature draws a liquidity shock ζ_t, reducing the firm's cash on hand to $C_t^{\text{eff}} = C_t - \zeta_t$. The firm then chooses and pays $\chi_t(B_t + w)$ and services mandatory debt amortization δD_t, using cash C_t^{eff} first and, if needed, new borrowing ℓ_t with $0 \leq \ell_t \leq L_t$. At the end of pay cycle t, any unpaid portion of the contractual obligation becomes next period's arrears. Thus, post-payment arrears carried into cycle $t+1$ are defined as $B_{t+1} = (1 - \chi_t)(B_t + w)$. If $B_{t+1} > 0$ the firm incurs a reputational (or moral) fixed cost m of withholding wages. The worker observes B_{t+1} and chooses effort e_t, incurring a convex cost $\psi(e_t)$, to maximize utility given B_{t+1} and beliefs about next period's payment χ_{t+1}. Continuation beliefs depend only on B_{t+1}. Output $y_t = Af(e_t)$ is then realized, and both revenue y_t and any remaining cash C_t^{res} are carried forward as C_{t+1}. The debt stock evolves as D_{t+1} reflecting new borrowing ℓ_t and accrued interest. The employee-employer match survives to the next cycle with probability $p(e_t)$. If the match breaks, the firm faces a repayment obligation of outstanding wages, occurring with probability ϕ and scaled by a penalty factor $\xi > 1$. The firm is matched with an identical worker in the next period after paying hiring cost H. All parties apply a common discount factor β.

Workers Beliefs.

Assumption M9 (Workers' Expectation)**.** *Liquidity ζ_t is i.i.d. and independent of firm type θ. Observing χ_t does not lead workers to update a posterior over θ; continuation beliefs depend only on $B_{t+1} = (1-\chi_t)(B_t+w)$. Aggregation is written as the conditional expectation:*

$$\mathbb{E}\left[\, u\big(\chi^*_{t+1}\big(B_{t+1}, D_{t+1}, C_{t+1}; \Theta, \zeta_{t+1}\big)\, (B_{t+1} + w) | B_{t+1}\big)\right]. \tag{18}$$

where the expectation is taken over (Θ, D, C, ζ) under a product measure and impose distributional assumptions for tractability.

The Firm Optimization. Per-period firm profits are defined as output minus all potential costs the firm has to pay (debt amortization, wages and the cost of wage withholding).

Formally, this is defined as

$$\pi_t(B_t, D_t, C_t, \chi_t; \theta) = y_t - \delta D_t - \chi_t(B_t + w) - m(B_{t+1}). \tag{19}$$

Given the worker's response $e_t^*(B_{t+1})$, a firm of type $\theta = (A, r)$ solves

$$\begin{aligned} V_f(B_t, D_t, C_t; \theta) = \mathbb{E}_\zeta\Bigg[\max_{\chi_t \in [0,1]} \Bigg\{ & \pi_t(B_t, D_t, C_t, \chi_t; \theta) \\ & + \beta \overbrace{p\big(e_t^*(B_{t+1})\big) V_f(B_{t+1}, D_{t+1}, C_{t+1}; \theta)}^{\text{Continuation value if match survives}} \\ & + \beta \underbrace{\big(1 - p\big(e_t^*(B_{t+1})\big)\big)}_{\text{Probability match breaks}} \Big[-\overbrace{\phi \xi B_{t+1}}^{\text{Expected fine}} - \underbrace{H}_{\text{Rehiring cost}} + \overbrace{\beta V_f\big(0, D_{t+1}, C_{t+1}; \theta\big)}^{\text{Value after rehiring}} \Big] \Bigg\} \Bigg] \end{aligned} \tag{20a}$$

subject to

$$\pi_t(B_t, D_t, C_t, \chi_t; \theta) = y_t - \delta D_t - \chi_t(B_t + w) - m(B_{t+1}) \tag{20b}$$

$$C_t^{\text{eff}} + \ell_t = \delta D_t + \chi_t(B_t + w) \tag{20c}$$

$$0 \le \ell_t \le L \tag{20d}$$

$$D_{t+1} = (1 + r)\,(1 - \delta) D_t + \ell_t \tag{20e}$$

$$0 \le D_{t+1} \le D^{\max} \tag{20f}$$

$$C_t^{\text{res}} = \max\{0, C_t^{\text{eff}} - \delta D_t - \chi_t(B_t + w)\} \tag{20g}$$

$$C_{t+1} = y_t + C_t^{\text{res}} \tag{20h}$$

The firm's value function is a discontinuous function because ℓ switches on at $\ell = 0$, and the borrowing cap may bind at $\ell = L$. Additionally, $m(B)$ also jumps at $B = 0$. The function is piecewise smooth in χ and optimal χ^* may be at corners or at piecewise interior points.

Worker Optimization. We can write the worker's problem as

$$\widetilde{V}_w(B_t;\chi_t) = \max_{e_t\in[0,1]} \Big\{ u\underbrace{\big(\chi_t^*(B_t + w)\big)}_{\text{observed payment}} - \psi(e_t) \\ + \beta\Big[p(e_t)\, V_w(B_{t+1}, U_{\text{pay}}) + (1-p(e_t))\,\big(V^{\text{out}} + \phi\, u(\xi B_{t+1})\big)\Big]\Big\} \tag{21}$$

The worker chooses effort to maximize its value function. Importantly, before observing χ the worker averages over firm policies so that we can write, using the definition of expected utility from equation 18

$$V_w(B_t, U_{\text{pay}}) = \max_{e\in[0,1]} \Big\{ \underbrace{U_{\text{pay}}(B_t)}_{\text{expected utility}} - \psi(e_t) \\ + \beta\Big[p(e_t)\, V_w(B_{t+1}) + (1-p(e_t))\,\big(V^{\text{out}} + \phi\, u(\xi B_{t+1})\big)\Big]\Big\}. \tag{22}$$

With probability $p(e_t)$ the match continues, yielding future wages; with probability $1-p(e_t)$ it breaks, and the worker receives their outside option plus potentially outstanding wages and penalties. The worker's first-order condition is

$$\psi'\big(e_t^*(B_{t+1})\big) = \beta\, p'\big(e_t^*(B_{t+1})\big)\, \underbrace{\big[V_w(B_{t+1}) - V^{\text{out}} - \phi\, u(\xi B_{t+1})\big]}_{Z(B_{t+1})}. \tag{23}$$

It immediately follows that $e_t^*(B_{t+1}) = 0$ if $Z(B_{t+1}) \leq 0$. I.e. if the worker has a sufficiently high outside option, or if enforcement mechanisms are sufficiently strong, the worker's optimal response is put in zero effort as soon as a wage payment is not honored by the employer. If the worker's value from keeping the match alive is sufficiently high, in relative terms, however, we can derive an expression for the optimal effort response depending on the level of arrears B_{t+1}. Differentiating the first-order condition with respect to B_{t+1} and rearranging shows that the overall sign is determined by $Z'(B_{t+1})$.

$$\frac{de_t^*}{dB_{t+1}} = \frac{\beta\, p'(e_t^*) Z'(B_{t+1})}{\psi''(e_t^*) - \beta\, p''(e_t^*)\, Z(B_{t+1})}. \tag{24}$$

Equilibrium. Parameters (L, δ, r) and the support of ζ ensure: (i) for each (B, D, C, θ) there exists a feasible $\chi \in [0, 1]$; (ii) under optimal policies, D and C remain bounded; (iii) u is continuous and strictly concave, p is continuous, increasing and strictly concave, $m(\cdot)$ is bounded and weakly increasing, and F_Θ, F_ζ have bounded support.

Definition 1 (Stationary equilibrium)**.** *A stationary equilibrium consists of* $(V_f, V_w, \chi^*, e^*, U_{\text{pay}})$ *such that:*

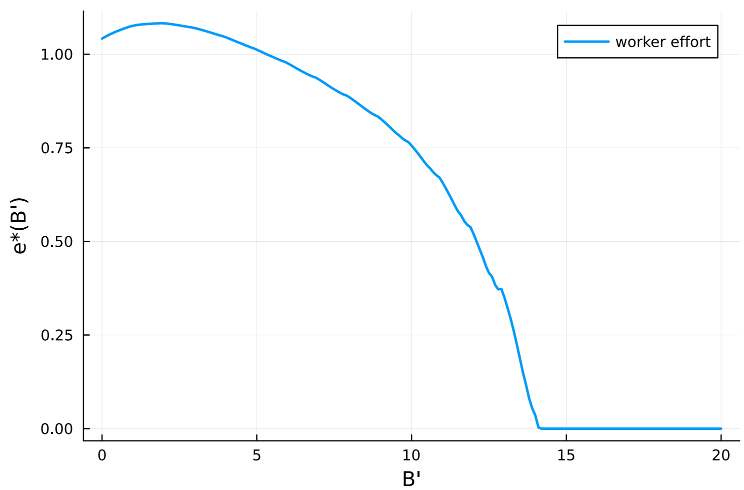
(i) ***Worker optimality:*** *Given* U_{pay}*,* V_w *solves* (22)*, and* e^* *satisfies* (23)*.*

(ii) ***Firm optimality:*** *Given* e^**, for each* θ*,* $\chi^*(\cdot)$ *solves* (20) *subject to* (20b)-(20h)*.*

(iii) ***Aggregation consistency:*** U_{pay} *is generated by* χ^* *via* (18)*.*

We solve for the equilibrium numerically using value function iteration. Figure F.1 illustrates the resulting equilibrium, showing optimal worker effort and aggregate firm policies for a parameterization calibrated to approximate the economic environment of Nigeria. Given reasonable low enforcement, we see that optimal effort can increase initially for low levels of arrears. At the same time, firms find it optimal not to pay workers with the repayment share declining as arrears grow.

Figure F.1: Worker Effort in Equilibriumn



NOTE: This Figure shows workers equilibrium effort. Workers beliefs about firm behavior is consistent with aggregate firm policy. Workers know the distribution of firm types in the economy and their respective optimal policy. Workers belief about the repayment probability is only determined via B_{t+1}, i.e. workers do not update beliefs about the firm type unlike in the main model without strategic interaction.

G Productivity Differences

G.1 Image Level Estimation

This Appendix presents averages corresponding to the daily estimates in Figure 7. We estimate the following modified version of equation (7) to obtain image-employee-level estimates:

$$y_{itq} = \alpha + \gamma\, R_i + \mu_t + \varepsilon_{itq}. \tag{25}$$

As before, y_{itq} denotes the outcome for individual i on date t for image q. R_i is a binary indicator for whether individual i was recruited in person, and μ_t are date fixed effects. We again exclude any individuals whose employment terms correspond to treatment arm 2 (uncertainty condition). We also estimate versions of equation (25) without date fixed effects and, respectively, with added image fixed effects. Even when averaging over the entire duration of the second employment round (before any salary delays occurred), we find no meaningful productivity difference between the in-person-recruited and job-advertisement-recruited samples. Results are shown in Table G.1.

G.2 Workday Level Estimation

This Appendix presents averages corresponding to the daily estimates in Figure 7. We estimate the following modified version of equation (7) to obtain employee-workday-level estimates:

$$y_{it} = \alpha + \gamma\, R_i + \mu_t + \varepsilon_{it}. \tag{26}$$

As before, y_{it} denotes the outcome for individual i on date t, R_i is a binary indicator for whether individual i was recruited in person, and μ_t are date fixed effects. We again exclude individuals whose employment terms correspond to treatment arm 2 (the uncertainty condition). We also estimate versions of equation (25) without date fixed effects. Aver-

Table G.1: **Average Productivity Differences (Image Level)**

	Continuous Index			Binary Index		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ In-person recruitees	0.026	0.025	0.034	0.001	0.001	0.001
	(0.035)	(0.035)	(0.032)	(0.006)	(0.006)	(0.006)
Mean (job-ad recruitees)	8.449	8.449	8.445	0.107	0.107	0.107
	(0.025)***	(0.025)***	(0.023)***	(0.004)***	(0.004)***	(0.004)***
R-Square	0.0006	0.0002	0.6288	0.0000	0.0000	0.7259
Observations	505,274	505,274	505,274	505,274	505,274	505,274
Individuals	221	221	221	221	221	221
SE Cluster	Individual	Individual	Individual	Individual	Individual	Individual
Date FE	Yes	No	Yes	Yes	No	Yes
Image FE	No	No	Yes	No	No	Yes

NOTE: This table reports average productivity differences between the in-person-recruited sample and the job-advertisement-recruited sample from the second employment round, estimated via Equation (25). Standard errors are clustered by employee. The sample excludes individuals who received employment terms corresponding to treatment arm 2. Columns (1)-(3) present results for the continuous productivity index, and Columns (4)-(6) present results for the binary index. The results indicate no meaningful productivity difference between the two samples.

aging over the entire duration of the second employment round—before any salary delays occurred—we find no meaningful productivity differences between the in-person-recruited and job-advertisement-recruited samples in workday-level outcomes. The only statistically significant difference concerns absenteeism: the in-person-recruited sample is slightly more likely to be absent. Results are reported in Table G.2.

Table G.2: Average Productivity Differences (Workday Level)

	Correct Images per Day		Hours Worked		Effective Hours Worked		Probability of Being Absent	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ In-person recruitees	-0.75	-0.70	-0.04	-0.04	-0.05	-0.04	0.04	0.04
	(2.33)	(2.33)	(0.06)	(0.06)	(0.04)	(0.04)	(0.01)**	(0.01)**
Mean (job-ad recruitees)	26.70	26.68	6.26	6.26	7.23	7.23	0.07	0.07
	(1.48)***	(1.48)***	(0.04)***	(0.04)***	(0.02)***	(0.02)***	(0.01)***	(0.01)***
R-Square	0.0004	0.0245	0.0004	0.1080	0.0007	0.0429	0.0042	0.0136
Observations	3,673	3,673	3,673	3,673	3,673	3,673	4,014	4,014
Individuals	223	223	223	223	223	223	223	223
SE Cluster	Individual	Individual	Individual	Individual	Individual	Individual	Individual	Individual
Date FE	No	Yes	No	Yes	No	Yes	No	Yes

NOTE: This table reports average productivity differences between the in-person-recruited sample and the job-advertisement-recruited sample from the second employment round, estimated via Equation (26). The results are estimated at the employee- workday level and standard errors are clustered by employee. The sample excludes individuals who received employment terms corresponding to treatment arm 2. Columns (1) and (2) present results for the number of correctly labelled images per workday. Columns (3) and (4) present results for the daily working hours inferred from an employee's first and last labelling submission of the day. Columns (5) and (6) present results for the daily effective working hours inferred from the timer that the *labelbox* software operates to track labelling time. Columns (7) and (8) present results for the probability that an employee is absent during the work period.

H Combining Estimates for Back-of-the-Envelope Calculations

H.1 Productivity Implications for Firms

We combine our intensive- and extensive-margin estimates to provide a back-of-the-envelope calculation of how much wage withholding affects firm productivity. Assuming a slack labor market—as in our setting—firms are not labor-constrained. We then compare observed productivity with a counterfactual scenario in which firms do not engage in wage withholding.

We use our estimated effect of salary reliability on job take-up (a 25 percent increase, from Table 5) as the share of workers that firms could replace with individuals who are currently unwilling to accept wage employment. For these marginal workers, we take the most favorable estimate of their relative productivity. Specifically, we use the upper bound of the 95 percent confidence interval around the point estimate in column (3) of Appendix Table G.1, which compares the productivity of job-advertisement recruitees and in-person recruitees. The upper bound suggests that in-person recruitees perform about 0.10 more correct labels per image on average.[71] Relative to a baseline of 8.4 correct labels per image, this implies a productivity difference of 1.15 percent.

At the same time, if firms do not engage in wage withholding, they forgo the productivity increase that withholding can generate. Using the lower bound of the 95 percent confidence interval from column (1) of Table 3, which gives the most conservative estimate of the productivity effect of delayed pay, implies a gain of only 0.014 additional correct labels per image.[72] Relative to the same baseline, this corresponds to a productivity increase of just 0.16 percent, which would no longer materialize.

Combining these effects, firms forgo the opportunity to replace 25 percent of their workforce with slightly more productive workers while also losing the productivity gains from

[71] $0.035 + 1.96 \times 0.032 = 0.097$
[72] $0.045 - 1.96 \times 0.016 = 0.014$

withholding pay:

$$0.25 \times 1.15 - 0.16 = 0.13. \tag{27}$$

Thus, wage withholding reduces firm productivity by roughly 0.13 percent in total. This estimate disregards other margins — such as absenteeism, which would further lower the estimate given that the in-person-recruited sample was slightly more absent. For expositional simplicity, we (generously) round this to 0.2 percent.

H.2 Welfare Implications and Efficiency Calculation

Our results suggest that wage withholding is socially inefficient due to its large costs for workers and their willingness to forgo a substantial portion of their salary for the certainty that wages arrive on time.

Even under the most favorable interpretation of the productivity effects we measure, this practice seems inefficient from a welfare perspective. Taking the upper bound of the 95 percent confidence interval for workers' effort response from column (2) of Table 3 implies that effort increases by approximately 12 percent when wages are delayed. Firms could achieve a comparable increase in effort through a standard performance-based incentive scheme. In our setting, this would correspond to offering a bonus of roughly 10 NGN (0.6 US cents) per correctly labeled image (see column (1) of Appendix Table I.1). Workers label on average around 250 images per day (see column (1) of Table 2), corresponding to about 5,000 per month assuming 20 full workdays. The baseline probability of labeling an image correctly is 10 percent (see Table 2), which rises to 11.2 percent with a 12 percent effort increase. Paying a 10 NGN bonus per correctly labeled image would thus cost around:

$$5,000 \times 0.112 \times 10\text{NGN} = 5,600\text{NGN} \tag{28}$$

(approximately USD 3.75) per month. While workers' willingness to pay for reliable salary delivery is about 22,500 NGN (USD 15) per month.

Hence, even under the most generous assumptions about the productivity effects of delayed wages, the practice is Pareto inefficient. Firms could instead reduce base pay modestly, offer simple performance-based incentives, and pay wages on time. Such a scheme would lower total wage costs while generating the same effort increase. At the same time, workers would benefit from reliable pay, and their willingness to pay for salary certainty would exceed the reduction in wages required to make this arrangement profitable for firms. In short, wage withholding is an inefficient equilibrium: both sides could be better off under timely payment and standard incentive contracts.

I Supplementary Data Collections and Interventions

This section describes the supplementary data collections and interventions.

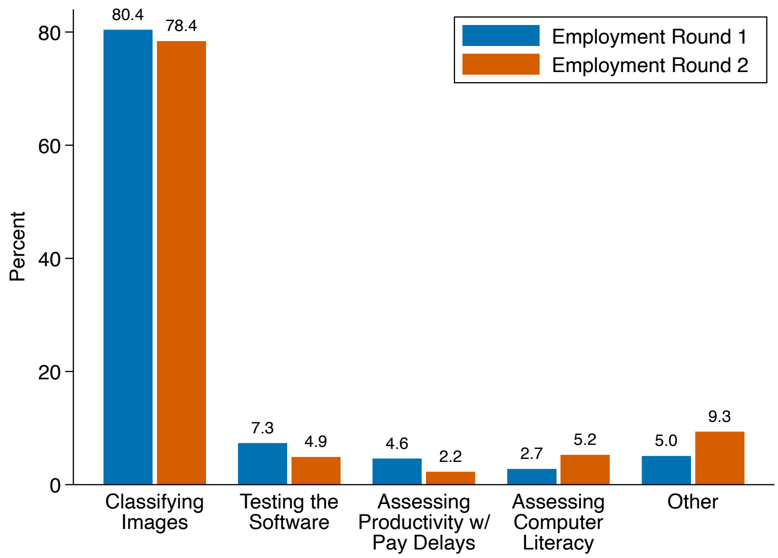
I.1 End-of-Job Survey

To elicit employees' beliefs about salary delays and their general perceptions of the job, we conducted an end-of-employment survey on the final workday of both employment rounds. The first survey was administered on May 2, 2025, the last workday of the first experimental employment round, and the second on October 3, 2025, the last workday of the second round. Both surveys were administered prior to the debriefing. We surveyed all employees present on those days—219 in the first round and 268 in the second. Surveys were administered individually to prevent responses from being influenced by peers, for example through social image concerns.

First, we asked employees about the purpose of the job. This was included to validate the experimental design and to ensure that employees perceived the job as genuine and primarily about image labeling. Responses are reported in Appendix Figure I.1 for both employment rounds. The figure shows that the vast majority of employees (80.4 percent in the first round and 78.4 percent in the second) believed the primary purpose of the job was image labeling. The remaining 19.6 percent and 21.6 percent, respectively, believed the job served another purpose, but only 4.6 percent and 2.2 percent, respectively, explicitly stated that it was to assess the productivity effects of salary delays. We interpret these results as evidence that our experimental setup was realistic and perceived as genuine.

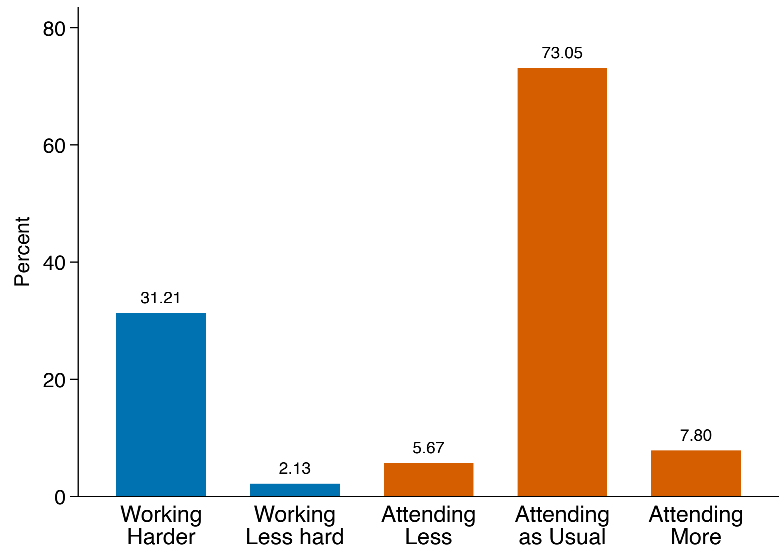
Second, we asked employees who experienced salary delays to self-assess how they responded to those delays. This was a multiple-choice question, and employees could select one or more options (conflicting responses were not possible). Results are reported in Appendix Figure I.2. Bars 1 and 2 refer to self-assessments of work effort: 31 percent of employees reported working harder when experiencing salary delays, while only 2 percent reported working less hard. Bars 3 to 5 refer to attendance: most employees (73 percent) indicated that they attended work as usual when their salaries were delayed, while small fractions reported attending more or less, respectively.

Figure I.1: Perceived Purpose of Employment



Note: This figure presents results from the end-of-employment survey conducted for both experimental employment rounds. The survey was administered to all 219 employees attending the final workday of the first employment round and all 268 employees attending the final workday of the second round. The figure shows employees' single-choice responses to the question: "In your opinion, what was the main purpose of this job and the program?"

Figure I.2: Self-Assessment of Responses to Salary Delays



NOTE: This figure presents results from an end-of-employment survey administered to 219 employees on the final workday. It shows employees' multiple-choice responses (conflicting responses were not possible) to the question: "How did you respond when your salary was delayed?"

I.2 Employment Terms Perception Survey

We conducted a follow-up survey with recruitees who were not hired to elicit their perceptions of the employment terms across the different treatment conditions. The survey assessed how treatments influenced recruitees' concerns about salary uncertainty. To avoid priming respondents about this issue, we framed the survey as a routine follow-up by the recruitment agency to evaluate the performance of the interviewer. A brief questionnaire was administered either by phone (by a different enumerator) or through automated WhatsApp messages. It included questions about recruitees' initial interview or information session, followed by their evaluation of the employment terms. We surveyed 204 individuals in total: 125 via WhatsApp and 79 via phone. Participants received a small monetary incentive in the form of mobile phone recharge cards for completing the survey. The administered questionnaire was identical between the phone and WhatsApp surveys. Respondents were asked how concerned they would be about receiving their salary *on time* and *at all* if they were to start the job. Participants had four response options *Very Concerned*, *Concerned*, *Little Concerned*, and *Not Concerned* corresponding to a four-point Likert scale.

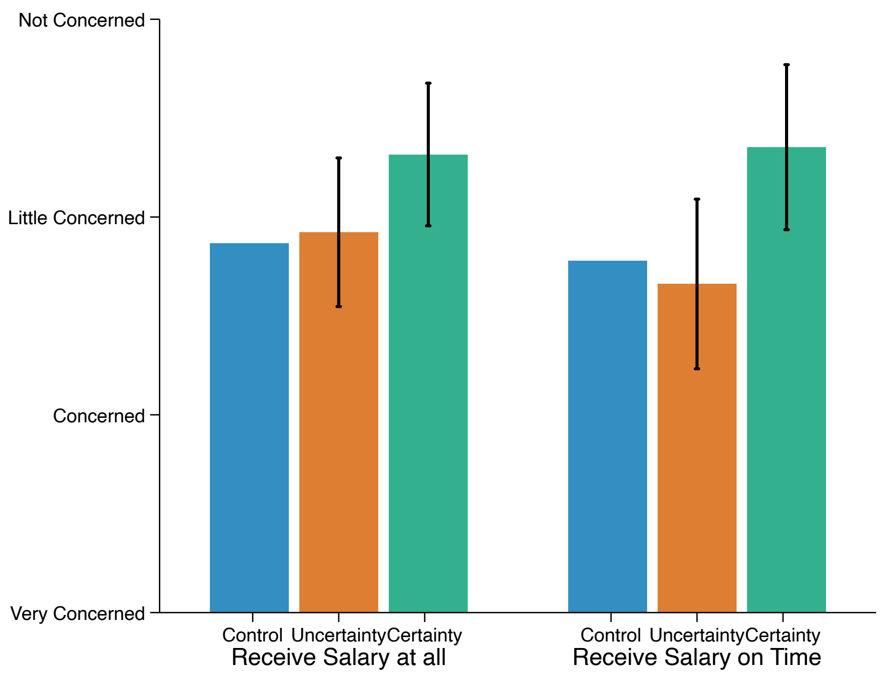
We then estimate the following regression:

$$y_i = \alpha + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \varepsilon_i. \tag{29}$$

where y_i is the numeric value of individual i's response, and Z_{ij} is an indicator for the type of employment terms j corresponding to one of the three treatment arms given to individual i. Treatment Arm 1 (Control Arm) serves as the omitted reference category.

Appendix Figure I.3 presents the survey results. The figure illustrates average levels of recruitees' concern about receiving their salary on time and, respectively, about receiving it at all. The patterns are fairly similar: individuals in the control arm report concern levels slightly above "Little Concerned." Those assigned to the uncertainty arm exhibit nearly identical levels of concern, which are not statistically distinguishable from the control group. This provides additional evidence of the high prevalence of wage withholding: telling people that salary may not be paid does not significantly change their concern about receiving their salary on time or at all. In contrast, the certainty treatment reduces concern. Recruitees offered employment terms with guaranteed salary are significantly less concerned than those in the control arm, reporting average levels of concern slightly below "Little Concerned."

Figure I.3: Salary Concerns by Treatment Arm (Follow-Up Survey)



Note: This figure presents results from the follow-up survey with recruitees on their perceptions of the treatment arms, estimated using Equation (29). Reported values are averages on a four-point Likert scale (1 = Very concerned, 2 = Concerned, 3 = Little concerned, 4 = Not concerned) and correspdond to the averages withing each treatment arm. We display 95% confidence intervals around the averages, based on heteroskedasticity-robust standard errors.

I.3 Bonus Payment

At the end of the first employment round—after all employees had received their outstanding salary balances—we implemented a bonus payment to contextualize the magnitude of our treatment effects. To estimate the causal effect of the bonus, we offered it to employees in selected work locations. The bonus amounted to 10 NGN per correctly classified image, corresponding to roughly two percent of the hourly minimum wage.[73] We then compare the performance of employees who were offered the bonus to those who were not by estimating the following regression:

$$y_{itq} = \alpha + \xi BP_i + \varepsilon_{itq}, \tag{30}$$

where BP_i is a binary indicator equal to one if employee i was offered the bonus payment. Our outcome variable, y_{itq}, is the binary productivity index, which is the appropriate measure here because the incentive applied only to entirely correctly labeled images. We additionally estimate versions of equation (30) in which we include data and image fixed effects. We restrict the estimation sample to the final week of the first employment round, during which the bonus was offered. Because bonus payments were provided only in some work locations, we can rule out potential spillover effects across employees. We find that offering the bonus increased the probability of labeling an image entirely correctly by 1.4 percentage points, corresponding to a 13.46 percent increase in performance relative to employees who were not offered the bonus. Results are reported in Table I.1.

[73] The minimum wage in Nigeria is defined as 70,000 NGN per month. Assuming 20 workdays per month and eight working hours per day yields an hourly wage of 437.5 NGN. To put this into perspective, this is equivalent to hiring a worker in the Illinois for the minimum wage of USD 15 per hour and offer a bonus of USD 0.3 per correctly completed task.

Table I.1: Effect of Bonus Payment on Labelling Performance

	Correctly labelled images		
	(1)	(2)	(3)
Effect of Bonus	0.014	0.014	0.012
	(0.007)**	(0.007)**	(0.007)*
Performance w/o bonus	0.104	0.104	0.105
	(0.006)***	(0.006)***	(0.006)***
R-Square	0.0004	0.0005	0.7889
Observations	290,181	290,181	288,718
Individuals	234	234	234
SE Cluster	Individual	Individual	Individual
Date FE	No	Yes	Yes
Image FE	No	No	Yes

NOTE: This table presents estimates of equation (30). We report the performance difference between employees who received a bonus and those who did not. The estimation is conducted at the employee-image level, and standard errors are clustered by employee. The sample is restricted to the final week of the first employment round, after all unpaid balances had been settled. The coefficient on the bonus payment implies that performance—measured as the probability of correctly labeling an image—increased by 1.4 percentage points (1.2 percentage points when image fixed effects are included). This effect is relative to a baseline probability of 10.4 percent (10.5 percent with image fixed effects).

J Compliers

In this section we discuss additional assumptions needed for the validity of the IV design.

Relevance. The relevance assumption requires that the instrument meaningfully affects the endogenous variable. In our setting, this means that receiving the high-certainty job offer (Treatment Arm 3) must shift individuals' attendance at the orientation day. This assumption is directly testable, and Table J.1 reports the corresponding first-stage regression. We obtain a strong first stage, with an F-statistic of 15.8.

Random assignment. Random assignment implies that the instrument is independent of potential outcomes. In our setting, this requires that the high-certainty job offer Z_3 is independent of individuals' potential attendance status at the orientation day, D_{id}. This condition holds by design, because participants were randomly assigned to treatment arms. Our final outcome variable is a predetermined characteristic of individual i interacted with attendance status. Therefore, if Z_3 is independent of potential attendance, it is also independent of the final outcome variable.

Exclusion. The exclusion restriction requires that the instrument affects the outcome only through its effect on the endogenous variable. In our setting, this means that assignment to the high-certainty salary condition can affect the outcome — the interaction of a predetermined individual characteristic with orientation-day attendance — only by influencing attendance at the orientation day. Because the individual characteristic is predetermined, the only channel through which the instrument can affect their interaction is by shifting attendance itself. Thus, the exclusion restriction holds in this setting.

Monotonicity. The monotonicity assumption requires that the instrument affects all individuals in the same direction. In our context, this means that receiving the high-certainty salary condition can only increase (and never decrease) an individual's likelihood of attending the orientation day. This assumption is plausible in this setting.

Treated and untreated compliers. By definition, the complier population consists of two groups: (i) those who receive the high-certainty salary condition and attend the orientation day, but would not have attended without the high-certainty offer; and (ii) those who do not receive the high-certainty offer and do not attend, but would have attended had they received it. These are the "treated" and "untreated" complier groups, respectively.

We can identify these groups using indicators for attendance and non-attendance, instrumented by indicators for receiving and not receiving the high-certainty offer Angrist et al. (2023). Under monotonicity, this approach isolates the two complier types. We can therefore estimate characteristics for each group. Our results pool these estimates and report the average across treated and untreated compliers.

Table J.1: First-Stage for Compliers

	(1)
Salary Guarantee	0.13***
	(0.03)
Observations	1,079
R-Square	0.45
F-Statistic	15.81
Date Control	Yes
Salary Control	Yes

NOTE: This table reports the first-stage estimates from equation (8). The dependent variable is attending the orientation day, and we estimate the effect of being assigned to the high-certainty salary condition.

K Choice Experiment

K.1 Design and Implementation

The choice experiment elicited recruitees' preferences over different job offers and was administered during the job interview for the sample responding to the job advertisement, respectively during the job information session for the in-person recruited sample. Recruitees were shown one of three choice blocks, each containing four binary choices between job offers. The job offers varied in two attributes: salary, and information on payment modalities. Salary was randomly varied across offers within the range of 55,000 to 85,000 NGN, while payment information varied according to the three treatment arms described in Section 4.1. Additionally, all job descriptions included the same information on location, hours, and job type.

Jobseekers were asked to indicate which of the shown job offers they would accept; possible responses included accepting one, both, or neither. We used a Bayesian D-efficient algorithm — a modified Fedorov algorithm (Cook and Nachtsheim, 1980; Zwerina et al., 1996; Carlsson and Martinsson, 2003) — implemented with the software developed by Hole (2015), to select the first three choices in each of the three choice blocks to maximize the statistical power of the design. Only the first three choices were used for estimating preferences, as specified in the pre-analysis plan. The fourth choice in each block was relevant for the branch of the experiment in which one of the respondent's choices could be implemented as an actual job offer, as described in the next paragraph.

Appendix Figure K.1 shows two example choices presented to recruitees. Panel (a) illustrates a choice between a job offer corresponding to treatment arm 3 (salary certainty arm) and treatment arm 2 (uncertainty arm). Panel (b) shows a choice between treatment arm 1 (control arm) and treatment arm 2 (uncertainty arm).

Most candidates were told that their choices would help us (the recruitment agency) offer job opportunities that matched their preferences. This constituted an incentive-compatible design for eliciting truthful responses, as candidates believed that the preferences they stated could directly influence the job offers they might receive. To further strengthen incentive compatibility, a subset of recruitees was randomly assigned to a different experimental condition in which they were told — truthfully — that one of their selected choices would be implemented as an actual job offer. While this approach is highly effective for eliciting truthful responses, the D-efficient design of the choice experiment limited the range of job-offer combinations we could feasibly implement. For this reason, we included a fourth choice in each block, which was not used in the preference estimation but provided additional flexibility for making offers in this subset of cases. The set of job offers we could extend remained constrained, which is why we implemented this highly incentive-compatible condition only for a subset of recruitees.

K.2 Theory and Estimation

To understand how employees value different job attributes and estimate WTP for higher salary certainty, we specify a discrete choice model.[74] Following the conceptual setup of McFadden (1974), we model individuals' utility from job choice j, denoted U_j, as deterministically dependent on observable job characteristics x_j and the w_j. Utility additionally depends stochastically on the unobservable term ε_j, so that utility U_j from job choice j can formally be expressed as

$$U_j = v(x_j, w_j; \beta) + \varepsilon_j, \tag{31}$$

where β is a vector of parameters. While utility also depends on the stochastic component, ε_j, it is deterministic from the perspective of the individual making the job choice. We further assume that individuals choose their utility maximizing job option j from the choice

[74] In this case, WTP is defined as the willingness to give up salary for greater salary certainty.

set C. The probability that an individual then chooses job j from the choice set C can be written as

$$\Pr(j|C) = \Pr(U_j > U_i) = \Pr(v_j + \varepsilon_j > v_i + \varepsilon_i) = \Pr(v_j - v_i > \varepsilon_i - \varepsilon_j)\ \forall\, i \in C. \tag{32}$$

To estimate this choice probability, we impose the standard restrictions on the structure of the utility function. First, we assume that ε_j follows a Type I extreme value distribution. Second, we assume that the utility function is linear in parameters. Third, we rely on the independence of irrelevant alternatives (IIA) assumption, which requires that the relative probability of choosing job j over i is unaffected by the inclusion or exclusion of other alternatives in the choice set. Under these assumptions the probability of choosing job j from the choice set C can be estimated using a conditional logit model. The probability of choosing job j as a function of the job characteristics x_l with $l = 1, \ldots, m$ and the salary w can then be written as

$$\Pr(j) = \frac{\exp(\sum_{l=1}^{m} \beta_l x_{jl} + \beta_w w_j)}{\sum_{k \in C} \exp(\sum_{l=1}^{m} \beta_l x_{jl} + \beta_w w_j)}. \tag{33}$$

We estimate the choice probabilities in equation (33) via maximum likelihood. Salary w is a continuous variable ranging from 55,000NGN to 85,000NGN. We include two additional variables capturing job attributes l. First, salary delay d is a continuous variable of the specified probability of salary delay between 0% and 50% (0% is the control condition). Second, the binary variable $s \in \{0, 1\}$ is a categorical variable indicating the usage of an automated payment system by the firm which is also reflected in the job offer. Standard errors are clustered at the individual level.

An advantage of the conditional logit model is that marginal rates of substitution (MRS) are straightforward to compute. This is particularly relevant in this case, because it allows us to derive a valuation of the different job attributes relative to salary. We interpret the MRS between a job characteristic and salary as a WTP for that job characteristic. The MRS between two attributes in the conditional logit model is simply the ratio of their estimated coefficients. Accordingly, the WTP for job characteristic l is given by:

$$WTP_l = \frac{\partial U / \partial x_l}{\partial U / \partial w} = \frac{\beta_l}{\beta_w} \tag{34}$$

K.3 Results

Table K.2 shows the full sample estimation results for the WTP estimates. Table K.3 replicates Table K.2 using a subsample only whose choice was implemented, making their answers highly incentive compatible. The tables show the MRS, i.e. coefficient ratio, between a job offer that offers higher salary certainty through the automated payment system and salary. The table also shows the MRS between the specified probability that salary delay may occur for a given employee and salary. Column (1) shows the estimation results using the entire sample. Column (2) shows the estimation results using the sample that responded to the job advertisement only, column (3) uses the in-person recruited and non-incentivized sample only while column (4) shows the estimation results for in-person recruited sample that had to be incentivized to participate in the job information session. The coefficient on 'WTP for Salary Certainty' in column (1) implies that jobseekers were willing to give up $27,223$ NGN ($\sim$ 18 USD) in monthly salary for a job offer that included the automated payment system and conveyed higher salary certainty. This is a substantial amount corresponding to 39% of the minimum wage of 70,000 NGN.

The coefficient on 'WTP for Delay' in column (1) implies that jobseekers are willing to accept a 1% higher probability that their salary will be delayed for a salary increase of 733 NGN.

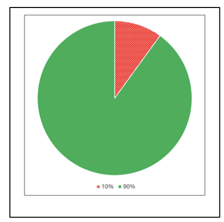
Table K.1: Attributes and Variation of Job Offers in the Choice Experiment

Job Attribute	Attribute Levels
Salary	55,000 NGN 60,000 NGN 65,000 NGN 70,000 NGN 75,000 NGN 80,000 NGN 85,000 NGN
Payment Mopdalitie	Biweekly Payments Biweekly Payments This company offers an automated payment system to guarantee on-time payment: your salary would be transferred automatically from a bank account with enough money to cover the salary payments. **No worker who received this has reported any issues with their salary.** Biweekly Payments Things in Nigeria are difficult at the moment, also for this company: every month there is a **{10, 20, 30, 40, 50} percent chance your salary will not be paid.**
Job Type	Image Classification
Location and Hours	Satellite Town, 9:00am to 5:00pm

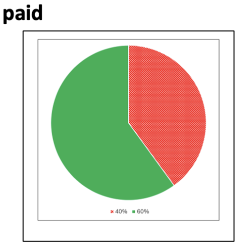
NOTE: This table shows the variation in job offers presented in the choice experiment. We vary salary and information on salary certainty according to the three treatment arms described in Section 4.1. All job offers provide the same information on job type, location, and hours.

Figure K.1: Examples of Job Offer Choices in the Choice Experiment

(a) Example Choice One

	Job Offer A	**Job Offer B**
Job Type	Image classification	Image classification
Location, hours	Satellite Town, 9:00am – 5:00pm	Satellite Town, 9:00am – 5:00pm
Monthly Salary	**55,000** NGN	**85,000** NGN
Payment	Biweekly This company offers an automated payment system to guarantee on-time payment: your salary would be transferred automatically from a bank account with enough money to cover the salary payments. **No worker who received this has reported any issues with their salary.**	Biweekly Things in Nigeria are difficult at the moment, also for this company: every month there is a **10% chance your salary will not be paid** 

(b) Example Choice Two

	Job Offer E	**Job Offer F**
Job Type	Image classification	Image classification
Location, hours	Satellite Town, 9:00am – 5:00pm	Satellite Town, 9:00am – 5:00pm
Monthly Salary	**55,000** NGN	**85,000** NGN
Payment	Biweekly	Biweekly Things in Nigeria are difficult at the moment, also for this company: every month there is a **40% chance your salary will not be paid** 

NOTE: This figure shows two example choices from the choice experiment as presented to recruitees.

Table K.2: Choice Experiment: Full Sample

	Entire Sample	Job Ad Sample	In-Person Sample (No Incentive)	In-Person Sample (Incentive)
WTP for Sal. Certainty	26.64***	25.99***	26.31***	28.37***
	(0.99)	(1.17)	(2.33)	(2.41)
WTP for Delay	-0.71***	-0.63***	-0.72***	-0.95***
	(0.02)	(0.03)	(0.05)	(0.06)
Observations	6,562	3,826	826	1,910
Individuals	1,110	641	139	330
SE Cluster	Individual	Individual	Individual	Individual

NOTE: This table shows the estimates of the choice experiment administered during the job interview. The table shows the coefficient ratios from a conditional Logit estimation, which can be interpreted as marginal rates of substitution. Standard errors are initially clustered at the individual level, and we calculate the standard error of the coefficient ratio using the delta method. Column (1) shows the estimation results using the entire sample. Column (2) shows the estimation results using the sample that responded to the job advertisement only and column (3) uses the in-person recruited sample only. Column (4) shows the estimation results for the sub-sample which was informed that one of their choices in the choice experiment would be implemented as their job offer.

Table K.3: Choice Experiment: Implemented Choices Subsample

	Entire Sample	Job Ad Sample	In-Person Sample (No Incentive)	In-Person Sample (Incentive)
WTP for Sal. Certainty	25.98***	25.11***	25.32***	28.54***
	(1.21)	(1.43)	(1.87)	(3.41)
WTP for Delay	-0.70***	-0.62***	-0.72***	-0.95***
	(0.03)	(0.04)	(0.05)	(0.08)
Observations	3,468	2,052	440	976
Individuals	586	344	74	168
SE Cluster	Individual	Individual	Individual	Individual

NOTE: This table shows the estimates of the choice experiment administered during the job interview. The table shows the coefficient ratios from a conditional Logit estimation, which can be interpreted as marginal rates of substitution. Standard errors are initially clustered at the individual level, and we calculate the standard error of the coefficient ratio using the delta method. Column (1) shows the estimation results using the entire sample. Column (2) shows the estimation results using the sample that responded to the job advertisement only and column (3) uses the in-person recruited sample only. Column (4) shows the estimation results for the sub-sample which was informed that one of their choices in the choice experiment would be implemented as their job offer.

K.4 Mixed Logit

In a second step, we relax the assumption of homogeneous preferences and estimate a random-coefficient (mixed) logit model that allows for individual-specific heterogeneity in tastes over job attributes (Train, 2003). We rely on Revelt and Train. (2000) to derive individual-level WTP estimates using maximum simulated likelihood.

Let n index individuals and j index job alternatives in the choice set C_n. Utility from job j for individual n is given by

$$U_{nj} = v(x_j, w_j; \beta_n) + \varepsilon_{nj}, \tag{35}$$

where the vector of taste parameters β_n now varies across individuals and is treated as a random draw from a continuous distribution $f(\beta \mid \theta)$ characterized by a finite-dimensional parameter vector θ (for example, the mean and covariance matrix of a multivariate normal distribution). Conditional on $\beta_n = \beta$, the choice probability takes the standard conditional logit form,

$$\Pr(j \mid \beta_n = \beta) = \frac{\exp\left(\sum_{l=1}^{m} \beta_l x_{jl} + \beta_w w_j\right)}{\sum_{k \in C_n} \exp\left(\sum_{l=1}^{m} \beta_l x_{kl} + \beta_w w_k\right)}. \tag{36}$$

The mixed logit choice probability for individual n is then obtained by integrating over the distribution of random coefficients,

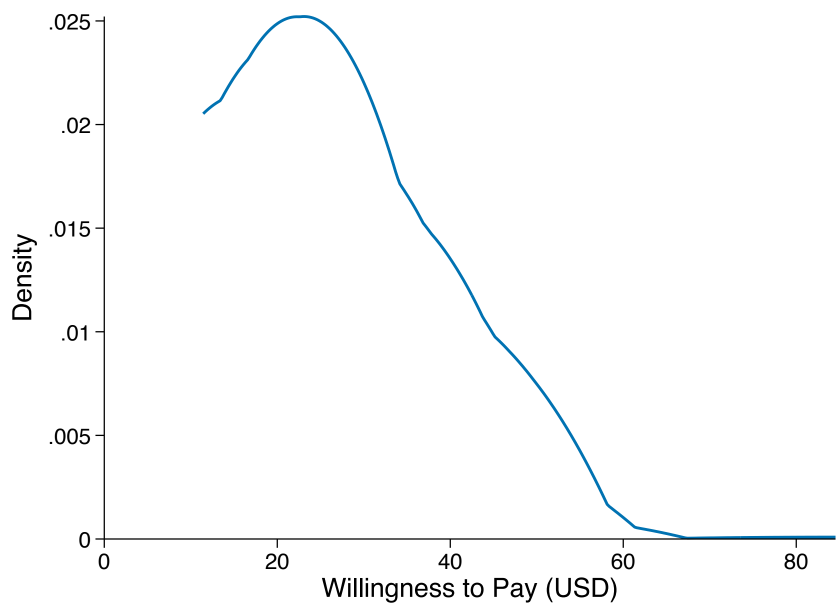
$$\Pr(j \mid C_n) = \int \Pr(j \mid \beta)\, f(\beta \mid \theta)\, d\beta. \tag{37}$$

Because the integral in equation (37) generally has no closed-form solution, we estimate θ by simulated maximum likelihood. The random-coefficient specification yields individual-specific draws (or posterior means) of the parameters β_n, and thus individual-level WTP measures. For individual n, the WTP for job characteristic l is

$$WTP_{l,n} = \frac{\partial U_n / \partial x_l}{\partial U_n / \partial w} = \frac{\beta_{l,n}}{\beta_{w,n}}, \tag{38}$$

which allows us to characterize the full distribution of WTP for salary certainty and other job attributes in the population.

Figure K.2: Distribution of Individual WTP for Salary Certainty



NOTE: This figure shows the distribution of individual-level WTP values, expressed in USD. These values represent the amount individuals are willing to forgo from their monthly salary in exchange for higher salary certainty. Individual-level WTP is estimated using equation (38).

Appendix References

Carlsson, Fredrik and Peter Martinsson, "Design techniques for stated preference methods in health economics," *Health Economics*, 2003, *12* (4), 281–294.

Chen, Angela, "Millions of Workers Are Training AI Models for Pennies," *Wired*, 2023. Accessed: September 5, 2025. Available at: https://www.wired.com/story/millions-of-workers-are-training-ai-models-for-pennies.

Cook, R. Dennis and Christopher J. Nachtsheim, "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics*, 1980, *22* (3), 315–324.

G2 Learning Hub, "20+ Data Labeling Statistics You Need to Know," 2023. Accessed: September 5, 2025. Available at: https://learn.g2.com/data-labeling-statistics.

Hole, Arne Risa, "DCREATE: Stata module to create efficient designs for discrete choice experiments," Statistical Software Components, Boston College Department of Economics Aug 2015.

Kässi, Otto, Vili Lehdonvirta, and Fabian Stephany, "How many online workers are there in the world? A data-driven assessment [version 4; peer review: 4 approved]," *Open Research Europe*, 2021, *1* (53).

Larousserie, David, "Humans Are Essential for Training AI," *Le Monde*, 2024. Accessed: September 5, 2025. Available at: https://www.lemonde.fr/en/science/article/2024/06/25/humans-are-essential-for-training-ai_6675713_10.html.

Market.us Insights, "Data Labeling Solution and Services Market News," 2024. Accessed: September 5, 2025. Available at: https://scoop.market.us/data-labeling-solution-and-services-market-news/.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," 1974.

Revelt, David and Kenneth Train., "Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier," Economics Working Papers E00-274, University of California at Berkeley 2000.

The Economist, "There Is a Vast Hidden Workforce Behind AI," *The Economist*, April 2025. Accessed: 31 October 2025. Available at: https://www.economist.com/international/2025/04/10/there-is-a-vast-hidden-workforce-behind-ai.

Vazquez-Bare, Gonzalo, "Identification and estimation of spillover effects in randomized experiments," *Journal of Econometrics*, 2023, *237* (1), 105237.

Zwerina, Katerina, Joel Huber, and Warren Kuhfeld, "A General Method for Constructing Efficient Choice Designs," Working Paper, Fuqua School of Business, Duke University 1996.