Gender and Financial Misconduct: A Field Experiment on Mobile Money

1 author:

Francis Annan
Georgia State University

GENDER AND FINANCIAL MISCONDUCT: A FIELD EXPERIMENT ON MOBILE MONEY

Francis Annan♣

Georgia State University

v: JANUARY 2022

Abstract

We construct a census of the market for mobile money in village Ghana and estimate that 1 out of every 4 mobile money transactions is overcharged relative to mandated rates. In an experiment, we randomize the matches between vendors and customers, finding strong evidence of "gender misconduct gap": female vendors are +37% more likely to commit such misconduct relative to male vendors. Misconduct is asymmetric: female customers are relatively more likely to suffer misconduct, and while female vendors discriminate against customers of their gender, male vendors favor their gender. Beliefs about gender, low female empowerment and income are relevant mechanisms.

KEYWORDS: *forensics and discrimination (**J16**, O12), household finance and fintech (**D18**, G23), culture and misconduct (**Z13**, G41)*

♣ Georgia State University, J. Mack Robinson College of Business, 35 Broad St NW, Atlanta, GA 30303. e-mail: `fannan@gsu.edu`. Field support from Samuel Adotevi, Kwamena Arkafra and Felix Debrah (all of Ghana Statistical Service) and all field officers is acknowledged. I thank, without implicating, Arun Chandrasekhar, Glenn Harrison, Paul Goldsmith-Pinkham, Erzo F.P. Luttmer, Vikas Agarwal, Stephen Shore, Alberto Chong, William Gerken, Liran Einav, Sandra Black, Jonathan Karpoff, Adriana Lleras-Muney, Xavier Gine, Belinda Archibong, Stephen Dimmock, Madeline McKelway, Sean Higgins (discussant), Justice Mensah and seminar participants for their helpful comments. I thank Advances with Field Experiments (University of Chicago), 3rd Toronto Fintech (University of Western Ontario), and NBER/RFS Inequality, Discrimination, and the Financial System 2021 conference participants for comments on an earlier draft of this paper. I am grateful to the World Bank for funding. The World Bank had no involvement in any substantive aspect of the research project. Institutional Review Board (IRB) approvals for research data collection were obtained from Georgia State University. The project was registered in the AEA RCT Registry, ID-0003812. This paper was previously titled "Gender and Financial Misconduct: Experimental Evidence from Mobile Money."

I Introduction

Misconduct – market actions that are unethical and indicative of fraud or wrongdoing – is a common and partially observed phenomenon that underlies many economic and financial transactions. Studies have begun to emphasize gender differences in financial misconduct, with large consequences for welfare (Egan, Matvos and Seru 2019; Annan 2020). Misconduct may lead to market discrimination if disproportionately committed against a particular gender. Similarly, it can lead to inefficient outcomes if misconduct reallocates resources from a more productive to a less productive gender group. Misconduct may also raise the marginal cost of transactions and decrease business activity if prices are perceived to be higher or uncertain, leading to inefficient outcomes. Such differences in gender and impacts are likely to be particularly important in settings with shallow formal institutions and where many people are arguably vulnerable and less financially sophisticated. Evaluating the sources and gender differences in misconduct is a significant yet poorly understood issue.

In this paper, we report on a study that examines the nature of misconduct in markets and quantifies its gender impacts. We draw on the local market for mobile money [M-Money] in Ghanaian villages. M-Money is an important financial market innovation in developing economies which has been shown to improve welfare and reduce poverty through a variety of causal channels (Suri and Jack 2016; BMGF 2021). It provides financial services and transactions which are delivered on digital mobile networks, and comprises market vendors, who are small business retail distribution outlets that provide cash-in and cash-out services to consumers (Human ATMs), earn transactional commissions as their profit, and exchange cash for so-called *e*-money i.e., electronic balances that can be sent from one account to another through SMS.

The market for M-Money provides a unique space to study gender and misconduct based on three appealing features. First, it is less regulated, compared to traditional banking and service providers have limited oversight into the behavior of market vendors. Second, it has the potential to disproportionately benefit very poor areas, where households or consumers have historically lacked access to formal banking, are arguably vulnerable, and are less financially sophisticated. The vast majority (95%) of localities have access and about 90% of households, their close family and friend networks have registered for a M-Money account. Third, the official charges on transactions are *ex-ante* set by providers that the market vendors work for, so vendors are not allowed to marginalize (Annan 2020). We use this feature to cleanly define misconduct as all transactions at the vendor point that are overcharged, which can be derived by comparing observed transaction charges to provider-approved prices. This form of misconduct is pervasive and can be found in many other countries. Recent consumer protection surveys of digital finance users show significant rate of vendor misconduct against consumers in Kenya (3%), Uganda (32%), and Nigeria (42%) (Blackmon, Mazer and Warren 2021).

In practice, studying gender aspects of misconduct on the market for M-Money, particularly in low-income environments, is challenging because relevant data on misconduct are unavailable, perhaps because it is difficult to detect and measure, and observed market transactions, if ever present, may suffer from market sorting which creates endogenous matches between market participants. This is a typical challenge that may confront studies using observed market data on transactions (Goldsmith-Pinkham and Shue 2019). Our research is designed to circumvent these potential challenges. First, we build a unique census of the market for M-Money across 166 poor and low-income communities in Eastern Ghana. We deployed trained field officers to visit each of these localities to list all vendors, and in some select localities also list all nearby customers who are within 5 houses radius around a given vendor; allowing us (i) to create a census of local markets which is defined to reflect the pair: vendor by the set of all nearby customers, and (ii) provide rich baseline information, general to specific, about the market.

Second, to cleanly measure misconduct and gender effects, we implement an innovative research approach that randomizes the matches between M-Money vendors and customers, and train customers to attempt standardized transactions with real monetary incentives. We use an extremely short and transparent transaction script. By using real transactions that span different transaction types, we recover rich information about market behavior and avoid major criticisms of standard audit studies within economics: deception and its subsequent effect on the market (Kessler, Low and Sullivan 2019). For identification, we exploit exogenous variation created by the random matches between customers and vendors. Our randomization ensures that customers are similar and by randomly assigning customers who are similar, our experiment eliminates endogenous matching between customers and market vendors to address concerns that customers select into vendors based on their own gender or the vendors gender.

We document three important findings. First, vendor misconduct is substantial: the overall incidence of misconduct is 27% and the average overcharged-amount due to misconduct reflects about 54-82% of mean official charges, thus imposing significant financial burden on households. Back-of-the-envelope calculations suggest that misconduct leads to an additional monthly cost burden of GHS20 *per* household, which aggregates to a monthly extra cost of GHS155 million (US$30.3 million) for the Ghanaian economy. Second, there is strong evidence of "gender misconduct gap": female vendors are 10 percentage points (pp) (equivalently +37%) more likely to commit misconduct relative to male vendors. Third, the nature of misconduct is asymmetric: female customers are 41-55% more likely to suffer misconduct relative to similar customers who are males. Relative to a male vendor-male customer match, female vendors are 20 pp more likely to cheat female customers but 13 pp more likely to cheat male customers. In contrast, male vendors are 15 pp more likely to cheat female customers relative a male vendor-male customer match. Interestingly, the former indicates evidence of within-gender discrimination, while the latter indicates within-gender favoritism. All market vendors, however, cheat female customers more relative to similar male customers. These effects are robust to several alternative model specifications, inference procedures, the influence of customers' fixed effect, and the use of post-double-selection LASSO for estimation

(Belloni et al. 2014).

Our results raise two main questions – (i) why are female customers overcharged more than male customers (gender discrimination) and (ii) why do female vendors overcharge more than male vendors (gender misconduct gap). We investigate several relevant hypotheses: beliefs about gender, low female empowerment, gender differences in vendors' household income, business size, market knowledge, shortfalls in business liquidity, and market transparency effects. For gender discrimination, we find that differences in beliefs about gender and low female empowerment are relevant explanations. We implement follow-up surveys that elicit first order beliefs (FOBs) and second order beliefs (SOBs) from market participants on various aspects of gender and misconduct, finding that vendors perceive female customers as less informed about finances. Similarly, the gender discrimination effects are stronger in environments where female customers are less empowered. For gender misconduct gap, we find evidence of income differences as a relevant channel. Female vendors have lower incomes relative to male vendors, and are more likely to cheat their customers because they are more dependent on the extra revenue from misconduct. We rule out alternative explanations such as non-intentional mistakes and differences in risk taking.

Misconduct and gender discrimination in payment markets is an open—and high priority—area of research, particularly in developing countries, where consumers lack experience with FinTech (Garz et al. 2021) and higher transaction fees can act as a barrier to the adoption of payment services (Higgins 2020; Annan 2020) and reduce risk sharing across households (Jack and Suri 2014). We make three contributions to the literature, with implications for policy. First, we add to the literature on financial misconduct (Karpoff and Lou 2010; Dimmock, Gerken and Graham 2018; Parsons, Sulaeman and Titman 2018; Egan, Matvos and Seru 2021). These studies have shown higher incidence of misconduct for males and explored variation in misconduct across space and in intensity. We show higher misconduct for females and complement this work by studying misconduct differences across gender lines. Despite the promise of M-Money (poverty reduction, risk sharing, resilience and personal finance, entrepreneurship impacts; see Bharadwaj, Jack and Suri 2019; BMGF 2021), to our knowledge, we are the first to document the amount and nature of misconduct using manipulated assignments of market participants from an emerging market setting.

Second, we contribute to the literature on market discrimination (see Bertrand and Duflo 2017 for a review). We complement this literature in two ways. Our evidence that for female customers “i.e., the marginalized”, the market for M-Money is an uneven playing field reaffirms previous work. The vast available evidence so far suggests that discrimination runs “across groups”, and not within-group (List 2004; Abbink and Harris 2019; Egan, Matvos and Seru 2021). We extend previous evidence and challenge theories of discrimination and matching to include “within-group” discrimination, based on our evidence that female vendors are more likely to cheat customers of their gender.

Third, we add to the literature on corruption in developing countries and forensics. Economists are often concerned with the question of “How much corruption or concealed behavior there is in developing countries?” (see Olken and Pande 2012 or Zitzewitz 2012 for surveys). Our market transactions and measures of misconduct, a form of corruption, provide a new estimate of potential corruption within a rural finance context, based on a new financial technology. We estimate a misconduct rate of 27% on incidence and 54-82% on severity or intensity, which fall within the range of estimates found in the corruption literature, although wide ranging. Our result on asymmetric misconduct illustrates that corruption may also be discriminatory with disproportionately negative effects on “vulnerable” customers (Hunt 2007).

From a policy perspective, increasing the share of females in organizations is often a common policy proposal for tackling market discrimination in finance. For example, in both developed and developing countries, there are initiatives that implement quotas for women on corporate boards. Pioneering examples include: in 2003, Norway obliged listed companies to reserve at least 40% of their director seats for women (Bertrand et al. 2019); in 2013, India mandated all listed companies to appoint at least one woman director on their boards. Our findings on within-gender discrimination contribute to these policy initiatives. We illustrate that such policies may not directly limit the misconduct gap or discrimination *per se* (Bertrand et al. 2019). Alternative policy steps, perhaps, will have to consider the underlying mechanisms.

The rest of the paper is structured as follows. Section II describes the experimental setting and data, the design, how we measure misconduct, and presents the basic descriptive evidence of misconduct. Section III presents our empirical strategy. Section IV documents the gender misconduct gap and asymmetry in misconduct on the market for M-Money. Section V explores the mechanisms. Section VI concludes.

II Setting and Research Design

II.1 Mobile Money

The market for M-Money is made up of vendors, customers, and service providers. M-Money vendors correspond to an outlet, shop, premises or local banking channels where M-Money transactions can be carried out on behalf of the providers – which are joint partnerships between mobile network operators (MNOs) and commercial banks. The vendors register accounts for customers and act as cash-in and cash-out transaction or banking points for customers. These vendors generically earn commissions on transactions by acting on behalf of the financial service operators. The introduction and significant penetration of digital mobile telecommunications have provided a cheap infrastructure to make M-Money services accessible even to the poor and low-income societies. In these poor environments, formal financial institutions are shallow and largely absent (see Banerjee and Duflo [2007; 2011] for authoritative surveys about this), making M-Money a competitive financial option in low-income environments.

Similar to other banking and financial services, the business of M-Money likely faces fraud and misconduct, which could take different forms. In policy circles, regulators from the Bank of Ghana, for example, have expressed concerns about such potential market misconduct, yet there is very limited quantitative evidence on the extent of financial misconduct on M-Money. There are ongoing regulator and stakeholder discussions about eliminating emerging risks and recognizable fraud on M-Money and providing ultimate consumer confidence in mobile financial services. For instance, in Ghana, the MNOs and their partners have been charged to build more risk-resilient financial infrastructures.[1] Our study is designed to not only estimate financial misconduct at vendor retail distribution points, but to characterize its nature and new mechanisms that rationalize observed gender differences and asymmetries in misconduct. We do this in a rural context where the business of M-Money could have larger impacts, if well designed.

In the next section, we discuss a baseline market census that we conducted and provide stylized facts about the market for M-Money, reflecting the setting of our study.

II.2 Market Census and Market Facts

II.2.1 Market Census

Detailed vendor x customer data on M-Money is unavailable. So, we carried out two unique censuses of the market for M-Money in Eastern Ghana between February-March 2019 (Baseline I) and January-February 2021 (Baseline II), spanning 10 districts. Districts are made up of sub-administrative units called "localities" or villages. Eastern Ghana was chosen for its two attractive features: it covers an expansive number of villages, with potentially M-Money vendor sites, and our initial pilot works in February 2017 (Annan 2017) in other parts of this region suggest substantial levels of misconduct on the market for M-Money. Our census exercise documents the universe of all vendor points, and in the case of Baseline I, other surrounding households who are located within 5 houses radius around a given vendor.

[1] "We also want you [Mobile Network Operators] to make your service affordable, we also want you [Mobile Network Operators] to put in place systems to minimize or eliminate fraud if possible and we also want you [Mobile Network Operators] to give wonderful customer service to your customers as they come to your premises to transact business. We want your system to have what it takes, to give very good audit trail of transactions." -- Bank of Ghana's payments oversight office head Clarence Blay, speaking at a stakeholder conference titled Expanding Cashless Payments Through Mobile Wallet Transactions, 2014. Available at: https://www.peacefmonline.com/pages/business/finance/201408/210849.php?storyid=100&

To focus on low-income environments and to ensure the presence of at least a M-Money vendor point in the locality, where customers can engage with transactions, we begin by restricting attention to localities across the eastern belt that have a total population between 1000-20,000 people. We use a master gazetteer of localities kept by the Ghana Statistical Service. With this restriction, we arrive at a total of 137 localities across 9 districts for Baseline I (Figures A.1 and A.2 display the spatial coverage) and 38 localities in one large district for Baseline II, which we shall refer to as "local markets". Trained field officers were deployed to visit each of the selected localities to list all vendors and the nearby customers. In practice, we find that 130 out of the 137 localities for Baseline I and 36 out of the 38 localities for Baseline II had one or more M-Money center(s) after we undertook the baseline market census (implying a 95% success rate).

II.2.2 Market Facts

The baseline census we conducted solicited information from all market participants: both vendors and customers. We asked information on their basic demographics, poverty and assets, detailed market records on M-Money and non M-Money services, including general to specific knowledge about M-Money transactions. Additional household information on personal finance, debts, savings, shocks and investments were obtained from customers. Here, we will focus on data that are relevant to our study of financial misconduct. Detailed summaries and other patterns about the market are available upon request.

Table 2 shows the summary statistics for the market separately for Baseline I and II. Female vendorship is 39% (36% for Baseline II), meaning that these local markets are disproportionately made up of more male vendors. However, 62% of the potential customers are females; customers are generally more likely to be self-employed, married and older than vendors on M-Money. Approximately and strikingly, only 50% (37% for Baseline II) of the vendors have received formal training about the market for M-Money before joining the business (this number is not statistically different between female and male vendors; see Table A.1). The overwhelming majority (90% [SD=0.29]) of customers, their close family and friend networks have registered for a M-Money account (also called "wallet"). The vendors are slightly less poor compared to customers: several indicators that are suggestive of less poverty are higher for vendors, e.g., household heads ability to read in English, small family size, access to proper toilet facility and other tangible assets.[2]

We turn next to specific features of the market. With an average experience of 2 years in doing M-Money business, a vast majority (75% [SD=0.43] for Baseline I and 60% [SD=0.49] for Baseline II) of vendors operate as a joint venture, bundling this with other services.[3] The average daily sales per vendor is about GHS2,260 [US$442] (not statistically different between female and male vendors; see Table A.1). For Baseline II, this is about GHS2,788. With a sales commission of 1%, the average vendor will earn a daily profit of around GHS25. Thus, most of these vendors operate relatively small to medium size enterprises. The majority of households or customers use M-Money services than other alternative commercial financial services: 94% of customers are M-Money users, 80% are formal bank users, while just 9% are post-office users. This can be explained by the potential ease and lower charges of M-Money, difficulty in access and distance to nearby services: we estimate an average distance of approximately 61 meters to the closest M-Money site, while this distance is about 383 meters for post-offices. Recall that we surveyed nearby households. Finally, in Table A.1, we break down the data for vendors by gender, illustrating that female vendors compare quite well with male vendors on several relevant variables in the market census.

[2] Poverty estimates, formally: since our study focuses on M-Money in low-income and poor environments, we fielded questions in the baseline market census that allow us to directly examine poverty. We adapted a recently develop short-cut—yet rigorous, inexpensive, simple and transparent—measure of poverty called the "Simple Poverty Scorecard" (Schreiner 2015). We estimate an overall poverty rate of 10.0% for the market vendors and 14.0% for the households/ customers. Details about this poverty scoring methodology can be found here http://www.simplepovertyscorecard.com/GHA_2012_ENG.pdf.

[3] We identified joint venture services like: groceries and provisions, local medicine, multi TV installation, registration of SIM cards, phones and accessories, airtime recharge cards, mini-credit transfers, acting as agents for land and house sales, electricals and accessories, photocopying and typesetting, educational/online results checking, electric prepaid credit, among others.

II.3 Research Design and Timetable

We design an audit experiment where experimental customers were given cash to make actual transactions on M-Money. We take this approach for two reasons: credible data on misconduct is directly unavailable, and it allows us to manipulate the market match between vendors and customers which is crucial for our analysis; eliminates the potential effects of market sorting between vendors and customers. Our setup and transactions embody three unique features that are worth noting: there is a random match of market participants based on gender, actual cash payments are utilized, and it spans multiple transaction types which are common in the market (12 different transactions in total): cash-in (deposits), cash-out (withdrawals) and account opening transactions.

The first feature allows us to credibly study gender differentiated effects, gender discrimination and favoritism, while the second helps to circumvent potential concerns that may underly measures of financial misconduct or fraud based on survey responses (DeLiema et al. 2018). In later sections, we compare these two measurement approaches. The third feature sets up a useful benchmark for falsification checks in the empirical analysis: transactions vary based on their vulnerability to vendor misconduct, e.g., transactions that are classified as over-the-counter (OTCs) may be more vulnerable relative to those transactions that are not OTCs. Finally, to mimic the local market context and properly capture misconduct, we recruit and use local residents who can speak and act similarly as traditional customers will typically act.

II.3.1 Timetable

Table 1 shows the details and timeline of the study.[4] Overall, we deploy our experiment in two phases: I-II, which correspond to the set of localities in Baselines I and II respectively. Phase I was run in 2019 as a baseline pilot experiment using a total of 4 trained customers (2

[4] In an early draft version of this paper, we used transaction data from only parts of phase I, which excludes the data from phase II: main experiment and some parts of phase I. See Table 1.

males, 2 females) across 130 localities in 9 districts, yielding 942 successful total transactions (N=942). Phase II was run in 2021 as the follow-up main experiment using a large total of 40 trained customers (20 males, 20 females) across 36 localities in one large district and yielded 1,165 successful total transactions (N=1,165). Pooling the two experimental phases together, for more variation, we have a total of 44 auditors and 2,107 transactions (N^{pooled}=2,107) to study a two category gender identity.[5] In comparison with recent audit experiment work (see e.g., Mujcic and Frijters 2021), we use a large number of experimental customers with potentially more statistical power. For instance, Mujcic and Frijters (2021), used 29 testers to evaluate racial bias in the Australian bus rides marketplace for a 4 category race identity.

Table 1: **Study timeline**

	DATE	ACTIVITY
Part 1	February 2017	**Pre-Pilot:** Misconduct – incidence, correlates, design (Annan 2017)
Part 2	February-March 2019	**Baseline I:** Market census – detail market records (130 localities)
Part 3		**Phase I: Pilot Experiment**
	August 2019,	Sampling
		Auditors recruitment (through field partners, GSS)
		Experimental customers-vendor assignments and training
	September-October 2019	**Sample:** 4 Auditors (2 Males, 2 Females) +
		126-130 randomly select vendors across 126-130 localities; (N=942)
Part 4		**Follow up measurements:**
	October 2019	Risk preferences elicitation
	April-May 2020	Beliefs about gender and misconduct elicitation
Part 5	January-February 2021	**Baseline II:** Market census – detail market records (36 localities)
Part 6		**Phase II: Main Experiment**
	February 2021,	Sampling
		Auditors recruitment (through field partners, GSS)
		Experimental customers-vendor assignments and training
	March-April 2021	**Sample:** 40 Auditors (20 Males, 20 Females) +
		163 vendors across 36 localities; (N=1,165)

5

Before we conducted the larger experiment, one concern we had with the pilot was that the males and females could happen to differ on some other important trait besides gender because of the few auditors. If we plot the distribution of misconduct for each of the four pilot experiment auditors (4-person plots), we find systematic patterns specific to gender, indicating that the effect is truly from gender and not other traits (see Figures A.5 and A.6). This suggests that our baseline pilot results will scale up with a large number of auditors. This motivated our main experiment and the pooling of the two experimental phases: here, all cross-auditor heterogeneity would be balanced by gender and reflect more the true population differences between male and female M-Money customers.

II.3.2 Randomization Design

Assignment of Auditors: We used a stratified random assignment in both phases of the experiment: we take all male auditors and randomly assign them half and half to male and female vendors. Then we do the same for all female auditors. This process ensures equal random assignments to opposite and same gender. We next discuss the randomization details and balance separately for the main and pilot experiments.

Main Experiment [Randomization and Balance]: 40 experimental customers (20 males, 20 females) were assigned to all the 163 vendors (104 males, 59 females) across the 36 localities in Baseline II. About 4 vendors covering 4 localities were uniquely assigned to each auditor. Here, the large number of auditors allows us to conduct meaningful tests for randomization balance. To test whether the randomization of auditors to vendors was successful, we run regressions that compare auditor characteristics based on both their gender and assignments. Table A.5 reports the results, and shows strong evidence of covariate balance and thus successful randomization. First, male and female auditors are strongly similar: the average characteristics of male auditors are not different from the average characteristics of female customers. Second, the average characteristics of auditors assigned to male vendors are not different from the average characteristics of auditors assigned to female vendors. Third, the average characteristics of female auditors assigned to male vendors are not different from the average characteristics of female auditors assigned to male vendors. The same hold for male auditors.

Pilot Experiment [Randomization and Balance]: 4 experimental customers (2 males, 2 females) were assigned to 130 select sample of vendors (73 males, 57 females) across the 130 localities in Baseline I. About 32 vendors covering 32 localities were uniquely assigned to each auditor. Each locality has about 3 vendors. So, to maximize statistical power, we randomly select one vendor per locality for the pilot field transactional exercises, which we shall refer to as representative vendors. Here, we examine balance at two levels. First, to what extent are the pre-transaction random samples of vendors or select markets representative of the entire market population? Sample representativeness and identification requires that being a representative vendor is independent of any relevant market-level statistics. To test that these samples are comparable to the market population, we run regressions that compare outcomes of selected vendors or markets to those that were not selected in the *pre* transactional exercise period. We consider a number of different relevant outcomes, and show that both sides of the market show no observable differences across the two groups. Tables A.3 and A.4 report the results, where we find no difference across markets selected and those not-selected to be representative. Second, was the randomization of auditors to vendors was successful? Due to the few number of auditors in the pilot experiment, we run regressions that compare auditor characteristics based on only their assignments to the 130 representative vendors (Mujcic and Frijters 2021). Table A.5 reports the results, and shows strong evidence of covariate balance. The average characteristics of auditors assigned to male vendors are not different from the average characteristics of auditors assigned to female vendors. In addition, the average characteristics of female auditors assigned to male vendors are not different from the average characteristics of female auditors assigned to male vendors. The same hold for male auditors. Post-transactions, male (female) auditors carried out roughly 54% (46%) of the total successful audit transactions (p-value=0.208 for the difference). The same hold for vendor specific transactions, which is reassuring and consistent with the pre-transactions evidence of successful randomization of experimental customers. In examining the pilot experiment, we include observed vendor characteristics in our main estimations (as in Mujcic and Frijters 2021) to account for potential imbalance.

II.3.3 Auditors' Training and Transaction Approach

Field auditors were chosen from our research partner's pool of field officers who reside in our study area, compare well demographically to the population of customers and with experience in carrying out local M-Money transactions.[6] The auditors were trained to follow the same approach on how to interact with the vendors, particularly use uniform language at visits to vendors and covered the same set of transactions. They memorized and implemented a very simple and transparent transaction approach: `Good morning /afternoon /evening. I want to make a M-Money transaction [USE CODES: T1...T12]. Present necessary details: phone number and sender or recipient details. Thank you for your service.` The codes T1 to T12 denote 12 different transaction types. See details in Appendix D. Both male and female auditors were tasked to carry out the same types and sizes of transactions. Auditors were initially endowed with GHS5,000 each since they had to perform the same set of transactions. They received half of this initial endowment in cash (to begin their cash-in transactions) and the other half on their M-Money wallets (to begin their cash-out transactions). Over time and depending on the amount of money lost due to true transactional charges or misconduct at vendor retail distribution points, we replenish their endowments for the subsequent transactions. At the end of the experiment, we did a final verification of the data and then paid the experimental customers their field allowances from the remaining money.

We implemented several quality controls for the transactional exercises (as in Annan 2020). First, we set up a computer-adaptive data collection platform (called data HQ), which allowed us to track and verify the data in real time and space. Right after every visit, auditors complete a brief questionnaire about the transaction (see Table D.1 in Appendix D) using their Tablets and synchronize the data to our data HQ for immediate access and verification. The GPS coordinates of all transactions are traceable. Second, we pre-piloted

[6] A potential concern is that vendors cheat strangers (like the auditors) but not the local repeat customers that they know. This is not a major concern here for several reasons. First, it might be more risky to cheat strangers because they might be more informed, which is especially true in this market context with much imperfect information. This reduces that possibility of systematically cheating strangers. Second, in our market environment, Annan (2020) estimates that a very large share of the market transactions are conducted with customers who have no family and/or close relations: customers from our study area were shown the locality-level roster of all vendors and then asked to indicate where they last transacted at and how they are related to that vendor: 8.0% of transactions were between participants who are blood-related, 22.0% were between participants who are friends, and 70.0% are not related at all. Third, we vary the type of transactions and auditors conduct multiple or repeat transactions at a vendor point to mimic repeat customers.

the proposed audit approach in February 2017 (Annan 2017), which yielded similar patterns of misconduct (as noted in the Market Census section). Third, we include transaction types that are either easy or difficult for the vendor to overcharge, finding consistent evidence, as we discuss below. Thus, our proposed approach has the strengths of objectively measuring misconduct (unlike other survey-based measures of misconduct with potential misreporting; DeLiema et al. 2018), while avoiding deception and its later effect on the market unlike other standard audit studies (Kessler, Low and Sullivan 2019).

II.4 Measuring Misconduct and Descriptive Evidence

Our field transactional exercises cover the 130 representative vendors in the pilot experiment and 163 vendors in the main experiment. Auditors are uniquely assigned to vendors, but multiple transactions are performed at each M-Money center at random, as long as such transactional services are available at the vendor point. There are instances where customers are unable to make certain transactions for a variety of reasons, including unavailability of network to insufficient *e*-bank cash. With transaction-type fixed effects, as we do in the empirical analysis, such service interruptions will have limited impact on our results. A simple regression of this transaction shortfalls on an indicator for whether the vendor is a female or not suggests no gender differences. About 4-6 successful trips were made per auditor per day to their assigned vendors. Compared with a typical large daily number of customers at vendor points, we do not expect this to meaningfully alter vendors behavior.

II.4.1 Measuring Misconduct

In our market setting, vendor misconduct can take different shapes including manipulation of "provider-approved" prices, fake transactions, unauthorized access and disclosure of customers' bank accounts, to other actions that result in profits. For our purposes, we define misconduct to entail transactions that are over-charged by the vendor when compared to the regulator or provider-approved tariff rates (Annan 2020; Egan, Matvos and Seru 2021). Here, vendors have the room to over-charge transactions because most vendor-involved transactions are not automated, vendor records, which typically excludes the fees, cannot be objectively verified (vendors record transaction information in their personal diaries), customers lack knowledge, among others. A major advantage of our framework is that we are able to measure misconduct at granular levels using the transactional exercises: (i) across different types of transactions, (ii) the specific incidence of it (extensive margin) and (iii) severity/ amount overcharged as a result of the misconduct (intensive margin).

II.4.2 Descriptive Evidence

Tables A.8 and A.9 report the descriptive statistics of vendors' misconduct overall and across different transactional classes for the pilot and main experiment phases (the full distributions are provided in Figures A.3, A.4, 1 and 2, for additional reference). From phase I: pilot experiment transactions, the overall incidence of misconduct is 23% [SD=0.41], with the average amount overcharged due to misconduct being GHS3.32 [SD=1.59], which is high because it represents about $\frac{3.32}{4.03} \times 100 = 82\%$ of the average "official charge" for the transactional amounts used in the study.[7] In phase II: main experiment transactions, the overall incidence of misconduct is 27% [SD=0.44] and the average amount overcharged due to misconduct being GHS2.19 [SD=2.124] or $\frac{2.19}{4.03} \times 100 = 54\%$ of the average mandated charges. In Table A.6, we break misconduct down by gender, and show that its incidence is substantially higher for female vendors (28% [SD=0.45]) compared to their counterpart male vendors (19% [SD=0.39]). Turning to the misconduct outcome on severity, there are similar patterns: the average overcharged amount due to misconduct is slightly higher for female vendors (GHS3.35 for females; GHS3.31 for males). This evidence is consistent across the various transaction types. Table A.6 replicates these patterns for the phase II: main experiment transactions.

[7] All shown and described in Table A.6, our field market transactions are allowed to vary in sizes of GHS50 (small), GHS160 (medium) and GHS1,100 (large). Their official charges are GHS0.50, GHS1.60 and GHS10.00 respectively. Thus, the average official charge, pooling all the 3 varying transaction sizes, is approximately GHS4.03.

We highlight three major aspects of the descriptive evidence on misconduct. First, there is heterogeneity in misconduct levels across gender and the different groups of transactions. Misconduct is higher for female vendors and concentrated in OTC transactions, which by construct are more vulnerable to vendor misconduct. More importantly, misconduct is limited in non-OTC transactions and does not significantly vary by gender (8% [SD=0.27] for female vendors; 5% [SD=0.23] for male vendors). Second, misconduct is potentially "costly" to consumers or households. The average false charges due to misconduct reflect about 54-82% of mean official charges, which may impose additional financial burden on households. With misconduct, transactions charges are around 1.7% (instead of the official 1% for a typical transaction). The total value of monthly M-Money transactions in Ghana is GHS22,118 million. Equivalently, the total value of monthly M-Money transactions per a household in Ghana is about GHS2,800.[8] For households, this implies a monthly cost of GHS48 *per* household (GHS20 due to misconduct). For the economy of Ghana, this implies a monthly cost of GHS376 million (GHS155 million due to misconduct). These are large and significant transactions costs. We shall also explore the gender effects by the severity of misconduct.

Finally, it is useful to compare our measure of misconduct "truths" i.e., derived from actual market transactions, with the alternative subjective measures i.e., typically derived from survey responses (see e.g., DeLiema et al. 2018). In our Baseline I market census, we fielded questions that ask households (as in DeLiema et al. 2018) to recall and indicate if any of the following circumstances happened on their M-Money account recently (i.e., within the past 3 months): (Qa) `someone used or attempted to use their accounts without permission`, (Qb) `unknown callers asking for their account information`, (Qc) `they carried out an incorrect M-Money transaction (e.g., to a wrong person; to a scammer)`, or (Qd) `have ever been overcharged M-Money fees at cash centers`.

[8] From Bank of Ghana (https://www.bog.gov.gh/wp-content/uploads/2019/08/PAYMENT-SYSTEM-STATISTICS-Ist-Quarter-2019.-1.pdf), the total value of M-Money transactions is GHS66,356.41 million between January-March 2019. We divide this by 3 to get the monthly total value of transactions. Next, we divide this by the total number of households in Ghana (~8 million) to get the monthly total value of transactions *per* household.

We use these responses to derive three separate measures of the incidence of misconduct m. The first measure, which is standard in the literature, combines (Qa), (Qb) and (Qc)

$$m = \mathbf{1}\{\mathbf{1}[(Qa) = Yes] \text{ or } \mathbf{1}[(Qb) = Yes] \text{ or } \mathbf{1}[(Qc) = Yes]\},$$

which are indicators that capture whether or not the households experienced any of the selected circumstances. The second measure simply uses (Qd)

$$m = \mathbf{1}\{(Qc) = Yes\},$$

and the third measure combines (Qa) and (Qd)

$$m = \mathbf{1}\{\mathbf{1}[(Qa) = Yes] \text{ or } \mathbf{1}[(Qd) = Yes]\}.$$

Results are displayed in the bottom panel of Table 2, and suggest misconduct incidences of 58% [SD=0.49], 19% [SD=0.40] and 30% [SD=0.45], respectively. These are either under- or over-measured, if compared to the overall truth of 23% from phase I: pilot experiment transactions, suggesting that one should be cautious in measuring and using misconduct based on survey responses. If misconduct is used as an outcome variable, then the practical effects of such measurement errors may be less severe. Measuring misconduct from actual market transactions, as we do, may be the preferred option for many reasons, but one shortcoming is that, its measures may not reflect the space of all feasible market transactions.

III Empirical Strategy

III.1 Intuition

The intuition for our identification strategy is straightforward. We exploit exogenous variations created by the random matches between vendors and customers. The misconduct of female vendors may differ from male vendors across the randomly assigned customers since there are existing gender differences, e.g., empowerment, that could create differential incentives for misconduct. Discrimination against female customers, within-gender favoritism and across-gender discrimination also create different incentives to influence the misconduct of vendors. Randomization ensures that customers are similar. By randomly assigning customers who are similar, our experiment eliminates endogenous matching between customers and market vendors, including consumer search to address concerns that customers select into vendors based on their own gender or the vendors gender. It does not address potential differences in shop attributes by vendors' gender since we cannot randomize vendors. We explore such differences in vendors gender and shop attributes as potential mechanisms underlying our results.

III.2 Model Specification

Our baseline analyses take two approaches. Both approaches use a simple linear regression framework to account for potential differences across gender. We begin with a model linking changes in misconduct m_{ivtd} to the gender of the vendor, Vendor: Female$_i$

$$m_{ivtd} = \beta \text{Vendor: Female}_i + \mathbf{X}_i'\xi + \eta_v + \mu_{td} + \epsilon_{ivtd} \quad (1)$$

where i, v, t and d index a vendor, market district (phase I: pilot experiment) or market locality (phase II: main experiment), transaction type, and transaction date, respectively. Results will be reported for the main experiment, pilot experiment and the pooled experiment data. The dependent variable m_{ivtd} is a dummy variable indicating that vendor i committed misconduct for transaction t at date d. In a separate set of analyses, we define m_{ivtd} as the severity of misconduct, reflecting the magnitude of overcharge paid to the vendor as a result of misconduct.[9] The independent variable of interest Vendor: Female$_i$ is a dummy variable which indicates that the vendor is a female. As a result, β captures the relative effect when compared to Male$_i$, the omitted category. Our full specification includes district

[9] We have zero values in the misconduct amount / severity outcome. To account for this, we also report results using an inverse hyperbolic sine `asinh` transformation.

or locality fixed effect η_v, and transaction $\times$ date fixed effect μ_{td}. These fixed effects allow us to compare male and female vendors who do business in the same geographic market area, the same transaction type and at the same transaction date, and accounts for unobservable differences based on location, transaction or market cycles.

Our second set of analyses is similar but focuses on the mismatch in gender between vendors and customers, and their interactions with differences in misconduct. To evaluate potential discrimination, we estimate

$$m_{ivtd} = \beta \text{Customer Assignment: Female}_i + \mathbf{X}_i' \xi + \eta_v + \mu_{td} + \epsilon_{ivtd} \quad (2)$$

where d indexes the date of visit. This exploits the audit design and random matches between customers and vendors to evaluate whether more financial misconduct is conducted against females once you eliminate endogenous customer-vendor matches and have the male and female customers acting similarly. We evaluate the nature of misconduct using the following saturated model

$$\begin{aligned} m_{ivtd} = {} & \beta_1 \text{Female-Female}_i + \beta_2 \text{Female-Male}_i + \beta_3 \text{Male-Female}_i \quad (3) \\ & \ldots + \mathbf{X}_i' \xi + \eta_v + \mu_{td} + \epsilon_{ivtd} \end{aligned}$$

where d indexes the date of visit. Female-Female$_{id}$ is an indicator for a gender match between a female vendor and female customer in period d, Female-Male$_{id}$ is an indicator for a gender mismatch between a female vendor and male customer, and Male-Female$_{id}$ is an indicator for a gender mismatch between a male vendor and female customer. β_1, β_2 and β_3 capture the relative effect when compared to Male-Male Match$_{id}$, the omitted category. β_3 measures gender favoritism or discrimination by male vendors against female customers (since the omitted dummy is Male-Male$_{id}$). Similarly, we compare β_1 and β_2 to examine discrimination by female vendors against female customers.

We account for vendor level observables such as their demographics, various business or shop characteristics in the vector $\mathbf{X}_i'$, including auditor's (i.e., experimental customer's) gender. We take a theory-driven approach and use machine learning (specifically LASSO) to select what out of the long list of controls we should include. We do this using the post-double-selection LASSO technique of Belloni et al. (2014). In our second set of analyses, where the matches are random, the post-double-selection LASSO for estimating the impacts deals with potential covariate imbalance. However, when looking at the link between misconduct and vendor's gender, where potential differences in shop attributes by vendors' gender might exist, the post-double-selection LASSO procedure allows us to look at how the differences in vendor misconduct is affected or explained away by the characteristics that the post-double-selection LASSO selects. Thus, we achieve good estimation performance, in addition to minimizing researcher degrees of freedom and the possibility for p-hacking. Notably, our descriptives and baseline estimates are very close to those from the post-double-selection LASSO estimation.

All standard errors are clustered at the select vendor level (phase I: pilot experiment) or local market level (phase II: main experiment) to account for correlations of transactions within vendor or local market respectively (Cameron and Miller 2015). For the main experiment, we also cluster the errors at either the auditor level (Mujcic and Frijters 2021) or the vendor level, yielding the same inference. We discuss effects that contain useful economic information (i.e., looking at effect sign and effect size; Abadie 2020).

IV Results

IV.1 Gender Differences in Misconduct

Table 3 reports estimates from multiple specifications of Equation (1). Observations are at the vendor $\times$ transaction $\times$ date level. The baseline effects of gender on the incidence of misconduct, which is defined as a dummy variable indicating whether or not the vendor committed misconduct at date d, are shown in the left panel. Results on gender differences for the amount overcharged, which is defined to reflect the amount overcharged and paid to the vendor as a result of misconduct are contained in the right panel, for alternative estimates on the severity of misconduct. The indicator for female vendor, Vendor: Female is positive and statistically significant. This implies that female vendors are more likely to commit financial misconduct compared to their male counterparts. The estimated misconduct difference is about 10 pp. With an overall misconduct of 27%, the estimated difference corresponds to $\frac{0.10}{0.27} \times 100 = +37\%$ higher misconduct incidence for the female vendors. For the intensity outcome, the effect is similar and corresponds to +35% (i.e., $\frac{0.251\text{GHS}}{0.713\text{GHS}} \times 100$).

Conditional on the local market, and transaction $\times$ date fixed effects that soak up potential confounding variation, we interpret this as evidence of "gender misconduct gap". In Tables, B.1, B.2, and B.3, we replicate these baseline results separately using data from the pilot experiment and pooled experiments transactions.

IV.2 The Nature of Financial Misconduct

We consider both the random assignment of customers and (mis)match in gender between vendors and customers, and use this to evaluate general market discrimination and how vendors treat their own gender types in terms of misconduct.

IV.2.1 Evidence of Market Discrimination

Table 4 shows the results from Equation (2). As indicated, $\mathbf{X}_i'$, includes vendors' gender in columns (3) and (6). The indicator for Consumer Assignment: Female is large and significantly positive across all outcomes and specifications. From the double-post-selection LASSO, we estimate that vendors are about +41% (11 pp) more likely to cheat female customers as compared to similar customers who are males. This corresponds to +55% for the severity outcome, and provides strong evidence that more financial misconduct is committed against female customers once you eliminate endogenous customer-vendor matches. These effects are consistent and larger using data from the pilot experiment and pooled experiments transactions (see Tables B.1, B.2, and B.3). As expected, results from the pooled experiments transactions are more significant.

IV.2.2 Treatment of Own Gender

Here, the analysis involves some level of sub-sampling as we compare the different market matches in gender.

We report the results from alternative specifications of Equation (3) in Table 5. Relative to a Male-Male Match, female vendors are more likely (with an estimate of +20 pp) to cheat female customers but +13 pp more likely to cheat similar customers who are males. However, male vendors are 15 pp more likely to cheat female customers relative to the match between a male vendor and male customer.[10] As shown, these results are robust to the various model specifications. The Male-Female estimate is economically meaningful ($\frac{0.15}{0.27} \times 100 = +56\%$) and statistically significant. This provides significant evidence that male vendors discriminate against female customers (or alternatively, favor male customers compared to the female customers). Comparing the Female-Female and Female-Male results, we estimate about +7 pp (20 pp-13 pp, respectively) more misconduct of female vendors against female customers. This difference is economically large (about +26%) and statistically significant at conventional levels in some model specifications (e.g., p-value=0.090 in column (2)). Indeed, looking at the severity outcomes in columns (4)-(6), there is significant evidence that female vendors also discriminate against female customers relative to similar male customers. Tables B.7, B.8, and B.9 replicate these effects separately across the pilot experiment and pooled experiments transactions.

Together, and when combined with the evidence of general discrimination against females (see Table 4), our results point to misconduct asymmetry: within-gender favoritism for males

[10] Thus, vendor misconduct is systematically higher against female customers as compared to similar customers who are males regardless of the vendor's gender. This further suggests that our estimated difference in misconduct between male versus female experimental customers is driven by the customers' gender rather than the customers' interpersonal variability (e.g., one being more gullible than the other when transacting), which follows from our strong evidence of randomization balance and the fact that customers were trained to use the same transaction approach.

and within-gender discrimination for females. As we noted earlier, conventional policies aimed at limiting discrimination in organizations and financial markets by increasing the share of females may not directly apply given the evidence of within-gender discrimination for female vendors. Our inference is thus congruent with Bertrand et al. (2019), who show no discernible overall labor market impact on women in business following Norway's 2003 corporate policy obliging listed companies to reserve at least 40% of their director seats for women.

IV.2.3 Where Should Female Customers be Transacting?

Our evidence on asymmetric misconduct indicates that for female customers, the market for M-Money is an uneven playing field because all vendors, regardless of gender, are more likely to cheat female customers than male customers. This motivates the following two questions. First, where should the "vulnerable" female customers be transacting at? In addition, where should the male vendors be transacting at, if the level of misconduct suffered vary from female to male vendors? Based on our results and the feature that M-Money provides a homogenous financial service, female customers are likely better-off if they transact with male vendors. Similarly, male customers are equally better-off if they transact with male vendors.[11]

V Possible Mechanisms and Discussions

Our results raise two main questions – (i) why are female customers overcharged more than male customers and (ii) why do female vendors overcharge more than male vendors. The former question has an intuitive and straightforward set of potential explanations; less so for

[11] Indeed, if consumers are financially sophisticated, then these results will imply that all customers, regardless of gender, will "sort" on the male vendors. However, in practice, this may fail due to binding frictions, e.g., existing social ties and inertia making the switch across vendors costly. This evidence motivates a test for financial sophistication based on market misconduct. One can evaluate if "savvy" consumers anticipate financial misconduct and whether that helps in keeping prices closer to official or mandated levels.

the latter. We gather survey data on beliefs about gender and misconduct, data on female empowerment from the Demographic Health Survey (DHS), and detailed information from both experiments to explore relevant explanations. The results can help guide policy designs aimed at reducing misconduct and discrimination in markets.

V.1 Why are Female Customers Overcharged More?

To understand why female customers are more likely to suffer misconduct, we explore two candidate reasons – (i) people believing that female customers are less informed about finances (Bordalo et al. 2019), and/or (ii) female customers being less empowered.

V.1.1 Differences in Beliefs about Gender

If, for example, vendors perceive male customers as more sophisticated, relative to female customers, then we might expect more vendor misconduct against the female customers. To explore this possibility, we deployed a phone survey of 214 subjects (32 vendors and 182 customers) across 32 localities to gather first order beliefs (FOBs) and second order beliefs (SOBs) of market participants on various aspects of gender and misconduct on M-Money. For six statements, reflecting the gender-differentiated misconduct effects from our main analysis, the subjects were asked to indicate their belief (`Agree`/ `Disagree`; i.e., FOBs) and incentivized guess about the percentage of others (all vendors and customers in their locality) that will `Agree` to the statements (SOBs). Details about the statements are contained in Table C.1 of Appendix D. Survey results are summarized in Figures 3 and C.1.

There is strong evidence that subjects (both vendors and customers) believe that male customers are more financially sophisticated. Fifty-eight percent of the respondents say that the male customers are more savvy in transacting M-Money, relative to female customers and the respondents estimate that 57% of others in the local market will agree that male customers are more savvy. No significant gender differences in beliefs exist for either vendors or customers but female vendors have a significant higher view that female customers are more easily overcharged. These results are consistent with why both vendors overcharge female customers more than male customers.

V.1.2 Low Women Empowerment

If female customers are less empowered, then we might expect more vendor misconduct against the female customers. To explore this hypothesis, we draw on data about women empowerment from the most recent DHS. We adapt two common indices of women empowerment (DHS 2014). Our first measure uses the number of decisions that women participate in alone or jointly, whereby higher values reflect a greater sense of entitlement and a higher status of women. The second measure uses the total number of reasons for which a husband is justified to beat his wife, where a lower score reflects higher levels of women's control and empowerment. This allows us to classify our districts into low (below median) and high (above median) women empowered market areas. The two measures are strongly correlated and generate the same classification for our nine study districts. We examine the influence of women empowerment using a modified version of Equation (2)

$$m_{ivtd} = \gamma \text{Customer Assignment: Female}_i \times \text{Empowered}_v + \beta \text{Customer Assignment: Female}_i \\ \ldots + \mathbf{X}_i' \xi + \eta_v + \mu_{td} + \epsilon_{ivtd}$$

If females are highly (or equally) empowered as males, then one would expect the match or assignment of customers to vendors to generate less differences in misconduct effects based on the gender of customers. Table 6 shows results, and provides consistent evidence that the disproportionate cheat against female customers diminishes under equal or high women empowerment. Notice that η_v absorbs the direct effect of Empowered_v and that we make the plausible assumption that measured women-empowered districts contain female customers that are empowered and vice versa. Customer-level data on female empowerment is not available.

V.2 Why do Female Vendors Overcharge More?

To understand why female vendors are more likely to commit misconduct, we explore heterogeneity in relevant vendor-level characteristics. We focus here on vendor-level characteristics since the district or locality fixed effect η_v in Equation (1) accounts for market-level effects, e.g., competition, presence of a formal bank in the locality, market-level peer effects along gender lines, etc. Differences in vendor characteristics by gender form a plausible set from which to draw potential theories or hypotheses that could be at play.[12] We test for gender differences in five relevant vendor characteristics – (i) income (i.e., vendor's household income), (ii) business size (i.e., vendor's average sales, an indirect proxy for competition at the vendor-level), (iii) market knowledge (i.e., vendor's experience in doing M-Money business and/or formal education status), (iv) liquidity shortfalls (i.e., vendor's likelihood to decline transactions due to insufficient liquidity at hand during audit transactional visits), and (v) market transparency (i.e., vendor's likelihood to post the official tariffs during audit transactional visits). We implement a two-step approach. First, we regress each of these characteristics against an indicator for whether the vendor is a female, Vendor: Female$_i$. Second, for characteristics that are meaningful and significantly different by gender, we examine their direct effects on vendor misconduct, $m_{iv(td)}$.

Results are shown in Tables 7 and 8. Two of the vendor characteristics are significantly different by gender: income and tariff posting/transparency. Such differences might create incentives for differential vendor misconduct. Female vendors have lower household incomes and thus more likely to cheat their customers because they are more dependent on the extra revenue from misconduct. Similarly, female vendors are less likely to post the official tariffs at their retail distribution outlets and thus more easily and likely to cheat their customers because misconduct may be hidden and hard to detect by customers. Next, to confirm

[12] In a separate set of analysis (omitted), we examine relevant market level characteristics (in the spirit of the low women empowerment exercise), finding no significant differences in competition (Reuben et al. 2015) and market-level peer effects (Bursztyn et al. 2014; Dimmock, Gerken and Graham 2018) by vendor's gender.

these hypotheses, we test the direct effects of vendor's income and tariff posting behavior on misconduct (see Table 8, showing both unconditional and conditional effects). Income is significant and has large negative effect on vendor misconduct, but tariff posting is not significant. We conclude that lower incomes for female vendors is a relevant explanation for why female vendors might overcharge more.

V.3 Stock of Mechanisms

We have explored several hypotheses that could rationalize the estimated gender discrimination and gender misconduct gap. For gender discrimination, our evidence provides support for (i) beliefs about gender, whereby vendors perceive male customers as more financially sophisticated than female customers and (ii) low female empowerment, whereby vendors overcharge female customers more for being less empowered. For gender misconduct gap, we find evidence of income differences, whereby female vendors have lower incomes and are more likely to cheat their customers because they are more dependent on the extra revenue from misconduct. Next, we evaluate two other potential channels.

V.4 Evaluation of Alternative Hypotheses

V.4.1 Intentional Misconduct or Non-Intentional Errors?

Our results on misconduct imply that misconduct is intentional and indicative of some-level of corruption or cheating behavior. An alternative interpretation is that they might reflect non-intentional errors committed by M-Money vendors when helping customers transact. If this was the case, then female vendors might more often commit such "errors" perhaps because they are systematically different from male vendors including: vendors' level of specialization (i.e., the extent to which the vendor's core business is carrying out M-Money transactions versus running a grocery store and occasionally assisting customers with M-Money transactions), distraction (i.e., the extent to which the vendor is focused on assisting with a particular transaction versus also running a grocery store or watching small children at the same time), and other potentially relevant dimensions.

We show evidence that our results on misconduct are more consistent with intentional misconduct. First, if misconduct is non-intentional and reflects only errors, then we should see more of (if not equally) both under-charging and over-charging relative to the mandated rate. However, if observed transactional charges are skewed to above the mandated rates (i.e., overcharging), then it would more likely reflect intentional misconduct or corruption. Figure B.1 shows the distribution of actual transactional charges relative to the mandated rates. This measures the likelihood of undercharging (if the difference is negative), correct-charging (if the difference equal to 0), and overcharging (if the difference is positive). Substantively, the differences between observed charges and mandated rates are strictly bounded below at 0, suggesting that misconduct is intentional. Next, we note that our results on misconduct asymmetry do not support innocent errors; rather, these are more consistent with intentional misconduct. The finding that misconduct is likely intentional is also consistent with the heterogeneity analysis (Table 7) which shows no significant differences in several relevant vendor attributes.

V.4.2 Differences in Risk Attitudes?

In a follow-up exercise, we re-visited a representative subset of the vendors to elicit their risk aversion following Gneezy and Potters (1997). We do not find any meaningful gender differences in risk aversion (see Figure B.2). Thus, gender differences in risk taking (Croson and Gneezy 2009; Charness and Gneezy 2012) and non-intentional errors are not major underlying mechanisms. Although we find very limited support for other relevant explanations, it is interesting to explore these alternatives and compare them based on gender.

VI Conclusions

We design a field experiment to provide new insights about gender differences in misconduct, a significant yet insufficiently understood issue that underlies many economic and financial transactions. We document new evidence of substantial misconduct, gender misconduct gap, discrimination and asymmetry on the market for M-Money–a growing and well-celebrated example of FinTech in developing economies. Female vendors commit (+37%) more misconduct relative to their male counterparts. All market vendors cheat female customers (+41% to +55%) more compared to similar customers who are males. While female vendors discriminate against customers of their gender, male vendors favor customers of their gender.

From a policy perspective, two implications can be drawn, based on our analyses. First, when the market environment is poorly regulated (as is usually the case for emerging markets and new financial products), misconduct may be significant and discriminatory. Second, our results illustrate that beliefs about gender, low female empowerment, and gender differences in vendor income are relevant explanations for gender discrimination and misconduct gaps. We do not find support for several other possible mechanisms based on a plethora of tests. This implies that a specific form of social distance (beliefs about gender, low female empowerment, and low female vendor income) can lead to undesirable market outcomes and may be an important source of financial market frictions. Tackling these may provide an alternative policy step in limiting financial misconduct and discrimination in transactional markets. Together, our results will likely be relevant for other market settings where consumer sophistication is low, women empowerment and income is low, and financial technology is emerging; for example, other sub-Saharan African countries and the Global South. Designing relevant market and consumer protection policies could take into account these gender differences to ameliorate misconduct and vendor bias on consumers.

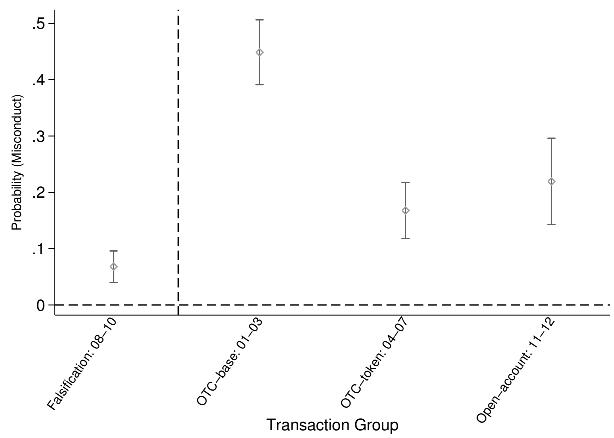
Our study provides an initial step towards the broader understanding of the nature and importance of misconduct in economic transactions, highlighting new and FinTech-based markets. Further research explores interventions that reduce misconduct and their market-wide impacts in the field. This line of work raises important issues at the intersection of economics and culture, with implications for the design of innovative financial instruments aimed at influencing financial market development and inclusion in low-income environments.

Table 2: **SUMMARY STATISTICS OF RELEVANT VARIABLES FROM THE MARKET CENSUS**

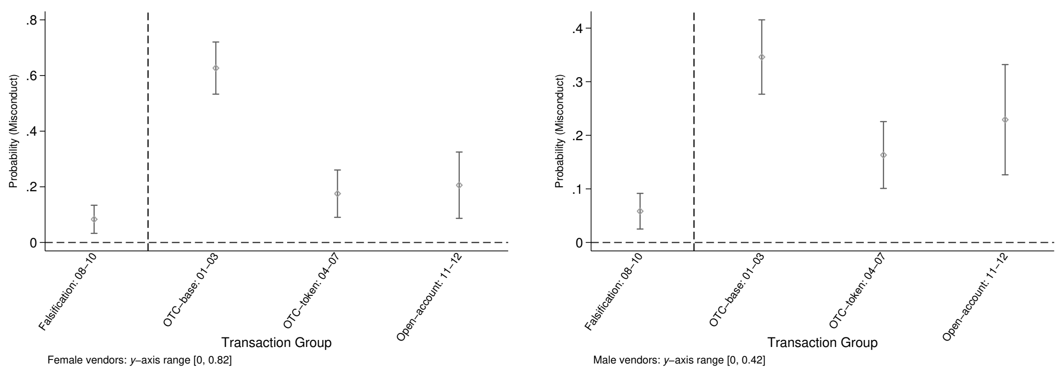
	Baseline I, 2019				Baseline II, 2021	
	Vendors		**Customers**		**Vendors**	
	Mean	SD	Mean	SD	Mean	SD
Female	0.39	0.489	0.62	0.484	0.36	0.482
Self employment	0.47	0.499	0.68	0.466	0.42	0.494
Self income intervals [GHS] (monthly)	2.01	1.483	1.37	0.868	2.10	1.325
Married	0.24	0.432	0.53	0.498	0.23	0.422
Akan ethnic	0.57	0.494	0.62	0.485	0.82	0.378
Age (years)	26.2	8.242	39.5	15.02	26.4	7.868
Education (any)	0.69	0.461	0.89	0.304	0.98	0.136
M-Money training	0.50	0.500			0.37	0.483
M-Money registered (self + any close person)			0.90	0.293		
Poverty Indicators						
Household size (above 5)	0.22	0.416	0.24	0.430	0.31	0.466
Household head read English	0.76	0.421	0.60	0.488	0.67	0.486
Outer wall used cement	0.74	0.433	0.70	0.456	0.822	0.383
Toilet facility	0.89	0.311	0.84	0.357	0.97	0.160
Working mobile phone(s)	0.97	0.152	0.97	0.151	1.0	0.00
Own working bicycle/ motor bicycle/ car	0.28	0.449	0.21	0.410	1.0	0.00
Market: Features + Transactions + Sales						
Doing business experience (years)	2.05	2.12			2.05	2.238
Joint venture: M-Money + other services	0.75	0.431			0.60	0.491
M-Money: Total volume [GHS] (daily)	2260	3775			2788	4117
Non M-Money: Number customers (daily)	32.7	47.06			50.5	55.68
Non M-Money: Total volume [GHS] (daily)	155	164.5			373	493.4
Distance to closest formal bank (meters)			338	751.3		
Distance to closest post office (meters)			382	250.7		
Distance to closest M-Money (meters)			61.2	94.92		
Formal bank user (of nearby banks)			0.80	0.395		
Post-office user (of nearby offices)			0.09	0.290		
M-Money user (of nearby vendors)			0.94	0.224		
M-Money: Total use volume [GHS] (weekly)			144	396.2		
Non M-Money: Number use (weekly)			2.27	14.76		
Non M-Money: Total use volume [GHS] (weekly)			44.7	505.1		
Borrowing + Savings						
Likelihood to borrow via M-Money (1-5 scale)			1.47	0.877		
Likelihood to save via M-Money (1-5 scale)			2.11	1.213		
Assessment: Fraud or Misconduct						
Attempted fraud experience (any)			0.58	0.492		
Ever over-charged			0.19	0.403		
Ever over-charged + unauthorized account use			0.29	0.455		
Number of observations	333		1,921		163	

Note: Table reports the summary statistics of relevant variables from our market census separately for both sides of the market: vendors versus customers. This include information about demographics, poverty indicators, and market outcomes, respectively. Customers' borrowing and savings behavior and their subjective assessment of market misconduct on M-Money are also shown. Baseline I census cover 333 vendors across 130 localities and 1,921 customers or households across a space of 137 localities in 9 districts. The exchange rate during Baseline I market census period is US$ 1.0 = GHS 5.12. Baseline II census cover 163 vendors across 36 localities in one large district (Atiwa).

Figure 1: **PHASE I – MISCONDUCT: DESCRIPTIVE STATISTICS BY GENDER BASED ON AUDIT TRANSACTIONAL EXERCISES**



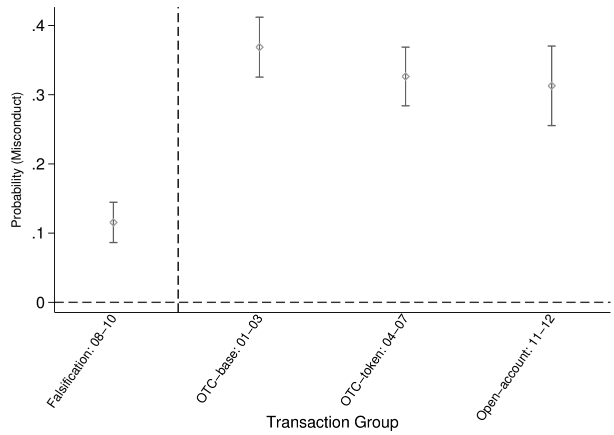
(a) **MISCONDUCT INCIDENCE × TRANSACTION GROUP**



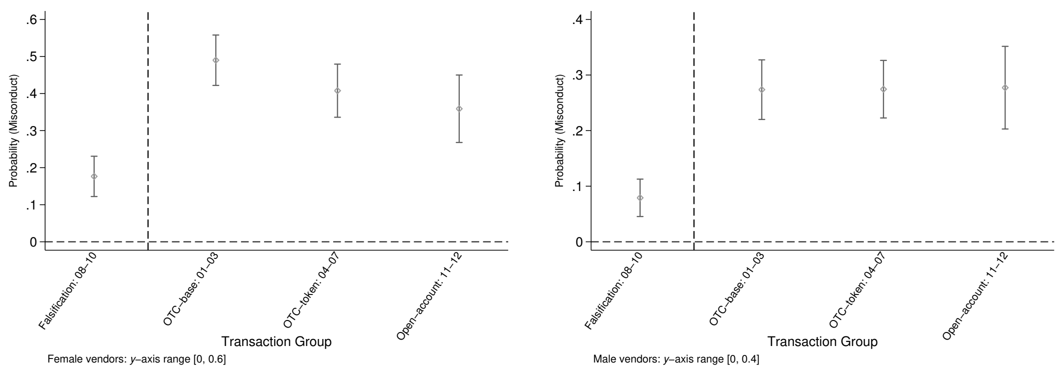
(b) **MISCONDUCT INCIDENCE × TRANSACTION GROUP × GENDER**

Note: Figures display the distribution of financial misconduct -- measured as the probability of the vendor committing a misconduct/ overcharging using actual transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation from the side of the customer. The specific transactions (01-12) in each transaction group are described in the Appendix, Table A.8. 90% confidence intervals are displayed around the estimates. Figure (a) shows the overall significance of misconduct and how it varies across the transaction groups. As expected, misconduct is much higher in the OTC-type transactions (i.e., little to no automation/verification required from the customer) compared to the Falsification group (automation and active verification required from the customer). Figure (b) shows how the overall significance of misconduct and how it varies across the transaction groups and gender. Misconduct is much higher in the OTC-type transactions compared to the Falsification group across gender. The probability of the vendor committing a misconduct is mostly higher for female vendors compared to male vendors.

Figure 2: **PHASE II – MISCONDUCT: DESCRIPTIVE STATISTICS BY GENDER BASED ON AUDIT TRANSACTIONAL EXERCISES**



(a) **MISCONDUCT INCIDENCE × TRANSACTION GROUP**



(b) **MISCONDUCT INCIDENCE × TRANSACTION GROUP × GENDER**

Note: Figures display the distribution of financial misconduct -- measured as the probability of the vendor committing a misconduct/ overcharging using actual transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation from the side of the customer. The specific transactions (01-12) in each transaction group are described in the Appendix, Table A.9. 90% confidence intervals are displayed around the estimates. Figure (a) shows the overall significance of misconduct and how it varies across the transaction groups. As expected, misconduct is much higher in the OTC-type transactions (i.e., little to no automation/verification required from the customer) compared to the Falsification group (automation and active verification required from the customer). Figure (b) shows how the overall significance of misconduct and how it varies across the transaction groups and gender. Misconduct is much higher in the OTC-type transactions compared to the Falsification group across gender. The probability of the vendor committing a misconduct is mostly higher for female vendors compared to male vendors.

Table 3: **Gender and misconduct gap**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Vendor:	0.138***	0.105**	0.101***	0.340**	0.187*	0.251**
Female (β)	(0.0325)	(0.0442)	(0.0311)	(0.127)	(0.111)	(0.110)
Observations	1,165	1,165	972	1,007	1,007	878
Locality FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.277	0.277	0.277	0.713	0.713	0.713
Number of localities	36	35	35	36	36	35
Number of auditors	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)

Note: Table shows the effects of vendors' gender on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual locality and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level over the period 3/2021-4/2021. Clustered standard errors (at the local market level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are similar if clustered at either the auditor level or the vendor level.

Table 4: **Gender and misconduct asymmetry – i**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Customer Assignment: Female (β)	0.0986**	0.100**	0.110**	0.582***	0.491***	0.383**
	(0.0403)	(0.0397)	(0.0495)	(0.121)	(0.127)	(0.179)
Observations	1,181	1,181	972	1,007	1,007	878
Locality FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.277	0.277	0.277	0.713	0.713	0.713
Number of localities	36	35	35	36	36	35
Number of auditors	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)

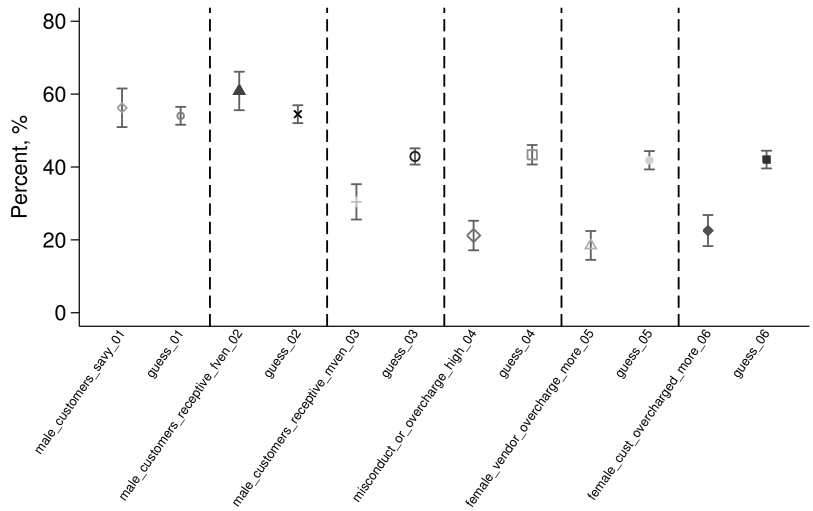
Note: Table shows the impacts of customers' gender assignment on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, 0-1 indicator for whether auditor is related to the vendor visited or not, and vendor's gender. The double-post LASSO specifications consider all vendor controls, and individual locality and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level over the period 3/2021-4/2021. Clustered standard errors (at the local market level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are similar if clustered at either the auditor level or the vendor level.

Table 5: **Gender and misconduct asymmetry – II**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Female vendor:	0.240***	0.174***	0.195***	0.892***	0.629***	0.568***
Female customer Match (β_1)	(0.0469)	(0.0521)	(0.0534)	(0.134)	(0.144)	(0.168)
Female vendor:	0.150***	0.0820	0.127*	0.240*	0.0332	0.0661
Male customer Match (β_2)	(0.0426)	(0.0599)	(0.0741)	(0.127)	(0.167)	(0.195)
Male vendor:	0.113**	0.0870	0.145**	0.510***	0.394**	0.416
Female customer Match (β_3)	(0.0483)	(0.0517)	(0.0636)	(0.153)	(0.190)	(0.265)
Observations	1,181	1,181	972	1,007	1,007	878
Locality FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	No	No	No	No	No
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.277	0.277	0.277	0.713	0.713	0.713
Number of localities	36	35	35	36	36	35
Number of auditors	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)
p-value (test: $\beta_1 = \beta_2$)	0.070	0.090	0.296	0.000	0.002	0.073
p-value (test: $\beta_1 = \beta_3$)	0.004	0.031	0.244	0.011	0.095	0.449

Note: Table shows the impacts of random gender matches between customers and vendors on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual locality and transaction × date fixed effects in the possible control set. Observations are at the vendor × transaction × date level over the period 3/2021-4/2021. Clustered standard errors (at the local market level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are similar if clustered at either the auditor level or the vendor level.

Figure 3: **Beliefs and incentivized guesses about gender and misconduct**



Note: Figure shows the distribution of market (32 vendors and 182 customers) beliefs across 6 selected statements about misconduct. The statements were designed to reflect the gender-differentiated market facts obtained from the main field trials. For each of the statements, market participants were asked to indicate their belief (i.e., `Agree`/ `Disagree`) and *incentivized* guess about the percentage of others (all vendors and customers in their locality) that will "`Agree`" to the statement: (01) `Male customers are more savvy financially`, (02) `Male customers are more receptive to female vendors overcharge behavior` (03) `Male customers more are receptive to male vendors overcharge behavior`, (04) `M-Money market misconduct or overcharging behavior is high`, (05) `Female vendors more likely overcharge customers`, and (06) `Female customers are more likely overcharged`, respectively. Details are contained in the Appendix, Table C.1. Each panel corresponds to a statement about misconduct. In each panel, the locality-level estimate of market belief (i.e., % of participants that `Agree`) is shown in the left, while the incentivized guess over the locality estimate is shown in the right. 90% confidence intervals are displayed around the estimates.

Table 6: **LOW WOMEN EMPOWERMENT AND MISCONDUCT ASYMMETRY**

	Incidence: Misconduct 0-1		Severity: Amount-Misconduct, GHS	
	(1)	(2)	(3)	(4)
Customer Assignment:	0.254***	0.222**	0.659***	0.669*
Female (β)	(0.058)	(0.091)	(0.179)	(0.382)
	[0.039]	[0.076]	[0.234]	[0.223]
x Empowered	-0.203***	-0.213*	-0.653***	-0.722***
	(0.057)	(0.127)	(0.156)	(0.166)
	[0.069]	[0.020]	[0.150]	[0.020]
Observations	942	936	867	861
District FE	No	Yes	No	Yes
Transaction x Date FE	Yes	Yes	Yes	Yes
Auditor FE	No	No	No	No
Controls	None	Double-Post LASSO	None	Double-Post LASSO
Mean of dependent variable	0.239	0.239	0.579	0.579
Number of districts (localities)	9 (126)	9 (126)	9 (126)	9 (126)
Number of auditors	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)

Note: Table shows the effects of women empowerment on the impacts of customers' gender assignment on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Empowered is a 0-1 indicator for localities in districts with higher (above median) women empowerment. Direct effect for Empowered soaked up in district FEs. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, 0-1 indicator for whether auditor is related to the vendor visited or not, and vendor's gender. The double-post LASSO specifications consider all vendor controls, and individual district and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level over the period 9/2019-10/2019. Clustered standard errors (at the select vendor level) are reported in parentheses. Clustered standard errors (at the district level) are also reported in brackets since we exploit a district-level variation. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level).

Table 7: **HETEROGENEITY: DIFFERENCES IN VENDOR CHARACTERISTICS**

	Income (1)	Business size (2)	Experience in business (yrs) (3)	Education **1**[> Primary] (4)	Illiquidity **1**[Decline transaction] (5)	Transparency **1**[Post tariff] (6)
Vendor:	-0.644**	1760	0.372	0.001	0.025	-0.143*
Female (β)	(0.311)	(4367)	(0.592)	(0.045)	(0.021)	(0.082)
Observations	68	129	129	139	1,836	
Locality FE	Yes	Yes	Yes	Yes	Yes	Yes
Transaction x Date FE	No	No	No	No	Yes	Yes
Mean of dependent variable	1.986	7586	1.684	0.974	0.396	0.402
Number of localities	29	35	35	36	36	36

Note: Table shows gender differences in relevant vendor characteristics. Vendor: Female is an indicator for whether the vendor is a female. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Income denotes vendor's household income. Vendors were asked to indicate their monthly income across five relevant income intervals: less than GHS500, [GHS501-GHS1,000], [GHS1,001-GHS1,500], [GHS1,501-GHS2,000] and above GHS2,000, which we convert to an ordinal scale of 1 to 5, respectively. Business size denotes total sales (GHS). Vendors were asked to indicate their sales made from M-Money business during the last month. Education is an indicator for whether the vendor attained more than primary school level of education. Illiquidity is an indicator for whether the vendor declined a transaction due to insufficient liquidity during the audit visits. Transparency is an indicator for whether the vendor posted the official tariffs at the business premise during the audit visits. Observations are either at vendor level for baseline measures (Columns (1)-(4)) or at vendor $\times$ transaction $\times$ date level for the audit measures over the period 3/2021-4/2021 (Columns (5)-(6)). Clustered standard errors (at the locality level) are in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level).

Table 8: **HETEROGENEITY: DIRECT EFFECTS OF SELECT CHARACTERISTICS ON MISCONDUCT**

	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
	(1)	(2)	(3)	(5)	(6)	(7)
Income	-0.037*		-0.041**	-0.121***		-0.109**
	(0.019)		(0.020)	(0.037)		(0.045)
Transparency: **1**[Post tariff]		0.050	-0.052		0.294	0.212
		(0.036)	(0.048)		(0.197)	(0.209)
Observations	500	1,165	496	436	1,007	436
Locality FE (Number of localities)	Yes (29)	Yes (36)	Yes (29)	Yes (29)	Yes (36)	Yes (29)
Mean of dependent variable	0.277	0.277	0.277	0.713	0.713	0.713

Note: Table shows the direct effects of select vendor characteristics on misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Income denotes vendor's household income. Transparency is an indicator for whether the vendor posted the official tariff sheet at the business premise during the audit visits. Observations are at the vendor level for the baseline measures (and $\times$ transaction $\times$ date level for the audit measures). Clustered standard errors (at the locality level) are in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level).

References

[1] Abadie, Alberto. 2020. "Statistical nonsignificance in empirical economics." American Economic Review: Insights, 2 (2): 193-208.

[2] Abbink, Klaus and Donna Harris. 2019. "In-group favouritism and out-group discrimination in naturally occurring groups." CSAE Working Paper WPS/2019-02.

[3] Annan, Francis. 2017. "Fraud on mobile financial markets: Evidence from a pilot audit study." NET Institute Working Paper No. 17-16.

[4] Annan, Francis. 2020. "Misconduct and reputation under imperfect information." Working Paper, Georgia State University.

[5] Banerjee, Abhijit and Esther Duflo. 2007. "The economic lives of the poor." Journal of Economic Perspectives, 21 (1): 141-68.

[6] Banerjee, Abhijit and Esther Duflo. 2011. *Poor economics: A radical rethinking of the way to fight global poverty*. New York: Public Affairs. 15th Edition.

[7] Belloni Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on treatment effects after selection among high-dimensional controls." Review of Economic Studies 81 (2): 608-650.

[8] Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." In Handbook of economic field experiments. Vol. 1 Elsevier pp. 309-393.

[9] Bertrand, Marianne, Sandra Black, Sissel Jensen, and Adriana Lleras-Muney. 2019. "Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway." Review of Economic Studies, 86 (1): 191-239.

[10] Bharadwaj, Prashant, William Jack and Tavneet Suri. 2019. "Fintech and household resilience to shocks: Evidence from digital loans in Kenya." Working Paper, MIT.

[11] BMGF. 2021. "Research brief: The impact of mobile money on poverty." Bill and Melinda Gates Foundation.

[12] Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli and Andrei Shleifer. 2019. "Beliefs about gender." American Economic Review, 109 (3): 739-773.

[13] Bursztyn, Leonardo, Florian Ederer, Bruno Ferman and Noam Yuchtman. 2014. "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions." Econometrica, 82 (4): 1273-1301.

[14] Cameron, Colin A. and Douglas L. Miller. 2015. "A practitioner's guide to cluster-robust inference." Journal of Human Resources, 50 (2): 317-373.

[15] Charness, Gary and Uri Gneezy. 2012. "Strong evidence for gender differences in risk taking." Journal of Economic Behavior and Organization, 83 (1): 50-58.

[16] Croson, Rachel and Uri Gneezy. 2009. "Gender differences in preferences." Journal of Economic Literature, 47 (2): 448-74.

[17] DeLiema, Marguerite, Martha Deevy, Annamaria Lusardi and Olivia S. Mitchell. 2018. "Financial fraud among older Americans: Evidence and implications." NBER Working Paper.

[18] Demographic and Health Survey [DHS]. 2014. Technical report, Ghana.

[19] Dimmock, Stephen, William C. Gerken and Nathaniel P. Graham. 2018. "Is fraud contagious? Coworker influence on misconduct by financial advisors." Journal of Finance, 73 (3): 1417-1450.

[20] Egan, Mark, Gregor Matvos and Amit Seru. 2019. "The market for financial adviser misconduct." Journal of Political Economy 127(1): 233-295.

[21] Egan, Mark, Gregor Matvos and Amit Seru. 2021. "When Harry fired Sally: The double standard in punishing misconduct." Journal of Political Economy (Forthcoming).

[22] Garz, Seth, Xavier Giné, Dean Karlan, Rafe Mazer, Caitlin Sanford, and Jonathan Zinman. 2021. "Consumer protection for financial inclusion in low and middle income countries: Bridging regulator and academic perspectives." NBER Working Paper No. 28262.

[23] Goldsmith-Pinkham, Paul and Kelly Shue. 2019 "The gender gap in housing returns." Working Paper, Yale University.

[24] Higgins, Sean. 2020. "Financial technology adoption." Revise and Resubmit, American Economic Review.

[25] Hunt, Jennifer. 2007. "How corruption hits people when they are down." Journal of Development Economics, 84 (2): 574-589.

[26] Jack, William and Tavneet Suri. 2014. "Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution." American Economic Review, 104 (1): 183-223.

[27] Karpoff, Jonathan M. and Lou Xiaoxia. 2010. "Short sellers and financial misconduct." Journal of Finance, 65 (5): 21879-1913.

[28] Kessler, Judd B., Corinne Low, and Colin D. Sullivan. 2019. "Incentivized resume rating: Eliciting employer preferences without deception." American Economic Review, 109 (11): 3713-44.

[29] List, John A. 2004. "The nature and extent of discrimination in the marketplace: Evidence from the field" Quarterly Journal of Economics, 119 (1): 49–89.

[30] Mujcic, Redzo, and Paul Frijters. 2021. "The colour of a free ride." Economic Journal (131), 634: 970-999

[31] Olken, Benjamin and Rohini Pande. 2012. "Corruption in developing countries." Annual Review of Economics, 4 (): 479-509.

[32] Parsons, Christopher A., Johan Sulaeman and Sheridan Titman. 2018. "The geography of financial misconduct." Journal of Finance, 73 (5): 2087-2137.

[33] Reuben, Ernesto, Paola Sapienza and Luigi Zingales. 2015. "Taste for competition and the gender gap among young business professionals." Technical Report, National Bureau of Economic Research.

[34] Schreiner, Mark. 2015. "Simple poverty scorecard--Poverty-assessment tool for Ghana." Available at: http://www.simplepovertyscorecard.com/GHA_2012_ENG.pdf

[35] Suri, Tavneet and William Jack. 2016. "The long-run poverty and gender impacts of mobile money." Science, 354 (6317): 1288–1292.

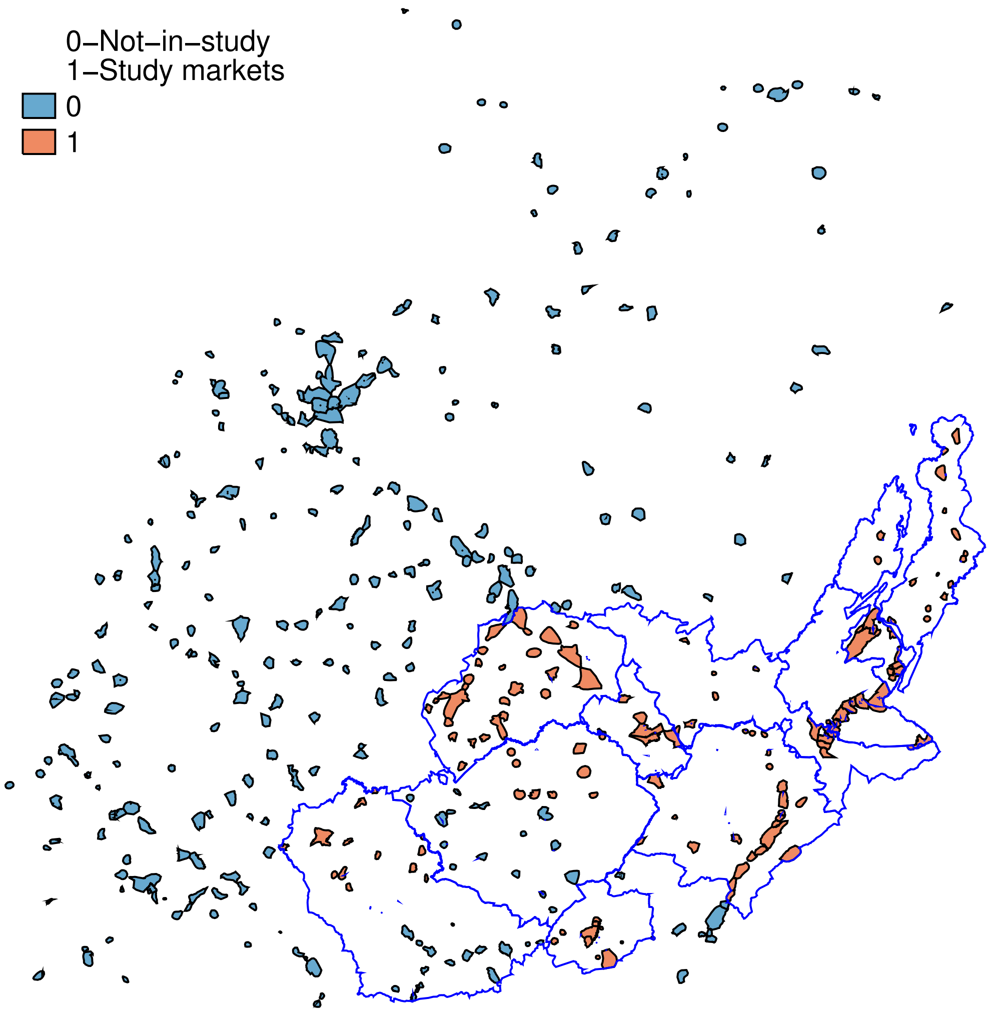
[36] Zitzewitz, Eric. 2012. "Forensic economics." Journal of Economic Literature, 50 (3): 731-69.

Supplementary Appendix

For Online Publication

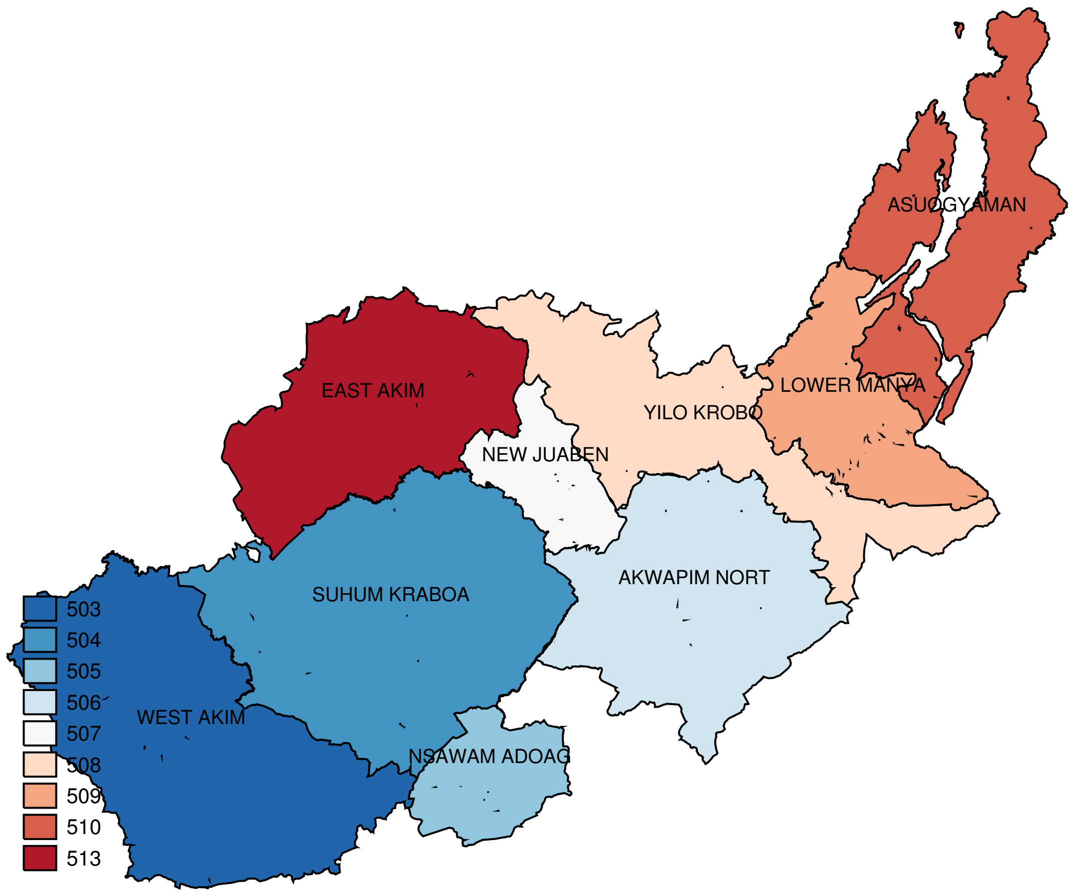
A Descriptive Statistics

Figure A.1: **BASELINE I – MARKET CENSUS: SPATIAL DISTRIBUTION OF THE LOCAL MARKETS**



Note: Figure shows the spatial distribution of localities in our study area (i.e., the eastern belt of Ghana). The polygons reflect localities. As displayed, 137 localities are selected for the Baseline I market census and subsequent experiment. Baseline I selected localities are located in 9 administrative districts, namely: West Akim, Nsawam Adoagyiri, Suhum Kraboa, East Akim, New Juaben, Akwiapim North, Yilo Krobo, Lower Manya Krobo, and Asuogyaman (district boundaries are displayed). To build the market censuses, we (initially) restrict attention to localities that have a total population between 1000-20,000 people to maximize the chance of having a M-Money vendor present in the locality.

Figure A.2: **BASELINE I – 9 EXPERIMENTAL DISTRICTS**



NOTE: The legend reflects district codes

Table A.1: **SUMMARY STATISTICS OF VENDORS BY GENDER FROM THE MARKET CENSUS**

	Baseline I, 2019				Baseline II, 2021			
	Females		**Males**		**Females**		**Males**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Self employment	0.43	0.496	0.50	0.500	0.27	0.450	0.49	0.502
Self income intervals [GHS] (monthly)	2.56	1.681	1.69	1.254	2.75	1.483	1.9	1.216
Married	0.27	0.445	0.23	0.424	0.24	0.431	0.22	0.418
Akan ethnic	0.57	0.495	0.57	0.494	0.88	0.326	0.79	0.403
Age (years)	25.7	7.823	26.6	8.493	26.5	9.082	26.4	7.126
Education (any)	0.74	0.435	0.65	0.475	0.96	0.184	0.99	0.099
M-Money training	0.52	0.499	0.49	0.500	0.31	0.466	0.39	0.491
Poverty Indicators								
Household size (above 5)	0.21	0.410	0.23	0.421	0.33	0.475	0.30	0.463
Household head read English	0.74	0.434	0.78	0.411	0.51	0.504	0.76	0.425
Outer wall used cement	0.74	0.435	0.75	0.432	0.81	0.392	0.82	0.380
Toilet facility	0.92	0.268	0.87	0.335	0.98	0.136	0.96	0.173
Working mobile phone(s)	1.00	0.000	0.96	0.195	1.0	0.00	1.0	0.00
Own working bicycle/ motor bicycle/ car	0.19	0.393	0.34	0.474	1.0	0.00	1.0	0.00
Market: Features + Transactions + Sales								
Doing business experience (years)	1.76	1.847	2.24	2.275	1.69	2.053	2.25	2.321
Joint venture: M-Money + other services	0.75	0.431	0.75	0.432	0.53	0.503	0.63	0.484
M-Money: Total volume [GHS] (daily)	2380	4927	2180	2757	2103	3652	3173	4326
Non M-Money: Number customers (daily)	26.1	26.47	37.2	56.35	53.1	66.47	49.46	52.25
Non M-Money: Total volume [GHS] (daily)	136	133.7	167	181.1	304	398.9	399.2	527.8
Number of observations	140		193		59		104	

Note: The exchange rate during Baseline I market census period is US$ 1.0 = GHS 5.12.

Table A.2: **PHASE II – BALANCE: AUDITOR ASSIGNMENTS TO VENDORS**

	Auditor Characteristics		Assignment: Pooled		Assignment: *if* Female vendor		Assignment: *if* Male vendor	
Characteristics of Auditors	Constant	**Female**	Constant	**Vendor: Female**	Constant	**Female**	Constant	**Female**
(Experimental Customers)	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Female	na	na	0.50***	0.008	na	na	na	na
			(0.049)	(0.081)				
Married	0.235***	-0.15	0.163***	-0.010	0.206***	-0.106	0.25***	-0.17**
	(0.085)	(0.120)	(0.036)	(0.060)	(0.067)	(0.094)	(0.050)	(0.071)
Akan ethnic	0.6***	0.05	0.615***	-0.022	0.586***	0.013	0.576***	0.076
	(0.110)	(0.156)	(0.048)	(0.080)	(0.092)	(0.130)	(0.067)	(0.096)
Age (years)	29.15***	-1.5	28.38***	-0.249	29.10***	-1.903	29.26***	-1.76***
	(0.728)	(1.029)	(0.316)	(0.526)	(0.548)	(0.769)	(0.444)	(0.628)
Education (post-college)	0.9***	-0.1	0.865***	-0.001	0.896***	-0.063	0.903***	-0.076
	(0.081)	(0.114)	0.033	(0.056)	(0.068)	(0.104)	(0.040)	(0.057)
Self employment	0.35***	-0.05	0.298***	-0.009	0.310***	-0.043	0.346***	-0.096
	(0.107)	(0.151)	(0.044)	(0.074)	(0.085)	(0.119)	(0.063)	(0.090)
Self income (1-5 scale)	1.55***	-0.25	1.403***	-0.047	1.448***	-0.181	1.557***	-0.307
	(0.288)	(0.407)	(0.126)	(0.210)	(0.247)	(0.346)	(0.176)	(0.249)
Has child	0.25***	-0.15	0.163***	0.006	0.241***	-0.141	0.25***	-0.173**
	(0.085)	(0.120)	(0.036)	(0.060)	(0.074)	(0.100)	(0.062)	(0.073)
M-Money Wallet	6.3***	-0.6	6.086***	-0.239	6.068***	-0.435	6.384***	-0.596
experience (years)	(0.414)	(0.585)	(0.182)	(0.303)	(0.345)	(0.483)	(0.257)	(0.363)
Household size	4.4***	0.05	4.567***	-0.075	4.379***	0.220	4.576***	-0.019
	(0.497)	(0.702)	(0.216)	(0.359)	(0.413)	(0.579)	(0.307)	(0.434)
Observations	40		163		59		104	
Joint F-test (linear), *p*-value	0.925		0.999		0.665		0.156	
Chi-squared test (probit), *p*-value	0.870		0.999		0.586		0.124	

Note: Observations are at the experimental customer level in column (1) and at the experimental customer × vendor pair level in columns (2)-(4). **Female** is a dummy variable indicating that the auditor is a female. **Vendor: Female** is a dummy variable indicating that the vendor is a female. The F and Chi-squared tests are conducted using all auditor characteristics. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A.3: **Phase I – Balance: pre transactions select-sample (customers)**

Demand side: Customers

	Constant	Select
Demographic Characteristics		
Female	0.62***	-0.002
	(0.022)	(0.026)
Married	0.51***	0.02
	(0.019)	(0.024)
Akan ethnic	0.62***	-0.002
	(0.036)	(0.039)
Age	38.63***	1.68*
	(0.737)	(0.891)
Education (any)	0.89***	0.009
	(0.015)	(0.016)
Self employment	0.66***	0.02
	(0.029)	(0.029)
M-Money registered	0.90***	0.001
	(0.014)	(0.017)
Poverty Indicators		
Household size	16.36***	-1.03*
	(0.508)	(0.559)
Household head read English	3.42***	-0.12
	(0.114)	(0.152)
Outer wall used cement	3.66***	-0.27
	(0.196)	(0.195)
Toilet facility	4.37***	-0.58
	(0.137)	(0.182)
Number working mobile phones	7.15***	-0.15
	(0.123)	(0.159)
Own working bicycle/ motor bicycle / car	1.18***	0.23
	(0.143)	(0.176)
Assessment: Fraud or Misconduct		
Attempted fraud experience (any)	0.61***	-0.04
	(0.040)	(0.039)
Ever over-charged/ unauthorized account use	0.29***	0.01
	(0.024)	(0.028)
Market: Features + Transactions		
Distance to closest formal bank (meters)	286.0***	147.8
	(73.10)	(107.3)
Distance to closest M-Money (meters)	66.29***	-10.75
	(12.78)	(13.021)
M-Money: Total use volume [GHS] (weekly)	129.2***	29.28
	(12.98)	(19.40)
Non M-Money: Number use (weekly)	2.062***	0.43
	(0.531)	(0.782)
Non M-Money: Total use volume [GHS] (weekly)	46.14*	-0.44
	(24.14)	(25.95)
Borrowing + Savings		
Likelihood to borrow via M-Money (1-5 scale)	1.515***	-0.06
	(0.073)	(0.069)
Likelihood to save via M-Money (1-5 scale)	2.12***	0.004
	(0.095)	(0.104)
Joint F-test (linear), *p*-value	0.181	
Chi-squared test (probit), *p*-value	0.206	

Note: Observations are at the customer level. The F and Chi-squared tests are conducted excluding all market outcomes. Standard errors (clustered at the locality level) are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A.4: **PHASE I – BALANCE: PRE TRANSACTIONS SELECT-SAMPLE (VENDORS)**

Supply side: Vendors

	Constant	Select
Demographic Characteristics		
Female	0.03***	0.02
	(0.049)	(0.076)
Married	0.20***	0.08
	(0.043)	(0.065)
Akan ethnic	0.57***	0.001
	(0.054)	(0.076)
Age	26.45***	0.71
	(0.585)	(1.117)
Education (any)	0.72***	-0.04
	(0.050)	(0.076)
Self employment	0.55***	-0.12*
	(0.058)	(0.075)
M-Money training	0.49***	0.04
	0.050	(0.070)
Poverty Indicators		
Household size	17.54***	-1.99
	(0.859)	(1.196)
Household head read English	4.10***	0.10
	(0.163)	(0.223)
Outer wall used cement	3.90***	-0.30
	(0.222)	(0.342)
Toilet facility	4.61***	-0.34
	(0.140)	(0.268)
Number working mobile phones	8.46***	0.36
	(0.208)	(0.261)
Own working bicycle/ motor bicycle / car	1.55***	0.71
	(0.287)	(0.499)
Market: Size + Sales		
M-Money: Total volume [GHS] (daily)	2296***	24.61
	(129.9)	(178.2)
Non M-Money: Number customers (daily)	32.82***	-0.02
	(1.796)	(2.520)
Non M-Money: Total volume [GHS] (daily)	156.4***	-0.72
	(6.272)	(8.799)
Joint F-test (linear), p-value	0.375	
Chi-squared test (probit), p-value	0.460	

Note: Observations are at the vendor level. The F and Chi-squared tests are conducted excluding all market outcomes. Standard errors (clustered at the locality level) are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A.5: **PHASE I – BALANCE: AUDITOR ASSIGNMENTS TO VENDORS**

	Assignment: Pooled		Assignment: *if* Female customer		Assignment: *if* Male customer	
Characteristics of Auditors (Experimental Customers)	Constant (1a)	**Vendor: Female** (1b)	Constant (2a)	**Vendor: Female** (2b)	Constant (3a)	**Vendor: Female** (3b)
Female	0.50***	-0.12				
	(0.061)	(0.098)				
Married	0.51***	-0.01	0.36***	-0.09	0.40***	0.06
	(0.061)	(0.100)	(0.103)	(0.136)	(0.090)	(0.131)
Akan ethnic	0.43***	0.08	0.26***	0.09	0.40***	0.06
	(0.061)	(0.100)	(0.087)	(0.136)	(0.090)	(0.131)
Age (years)	36.35***	1.17	0.70	-0.01	0.303	0.03
	(0.996)	(0.175)	(0.576)	(0.017)	(0.267)	(0.006)
Education (post-college)	0.51***	-0.015	0.36***	-0.09	0.40***	0.064
	(0.061)	(0.101)	(0.103)	(0.136)	(0.090)	(0.131)
Self employment	0.71***	0.11	0.26***	0.09	0.43***	0.00
	(0.056)	(0.082)	(0.087)	(0.136)	(0.065)	(0.000)
Self income (1-5 scale)	2.81***	-0.01	0.39***	-0.03	0.33***	0.03
	(0.156)	(0.252)	(0.141)	(0.045)	(0.204)	(0.065)
Has child	0.78***	0.01	0.36***	-0.09	0.43***	0.00
	(0.050)	(0.081)	(0.103)	(0.136)	(0.065)	(0.00)
M-Money Wallet experience (years)	7.71***	0.36	-0.29	0.09	0.43***	0.00
	(0.165)	(0.256)	(0.875)	(0.136)	(0.065)	(0.00)
Household size	4.74***	0.082	4.57***	-0.109	4.9***	0.13
	(0.099)	(0.173)	(0.087)	(0.156)	(0.174)	(0.267)
Joint F-test (linear), *p*-value	0.491		0.491		0.625	
Chi-squared test (probit), *p*-value	0.519		0.481		0.624	

Note: Observations are at the auditor (or experimental customer) × representative vendor pair level. Vendor: Female is a dummy variable indicating that the vendor is a female. There is no variation and thus no differences among auditors in the following additional auditor characteristics (0-1 indicators): has M-Money registered, household head read English, outer wall of house used cement, has toilet facility, own working mobile phone, and own working bicycle. The F and Chi-squared tests are conducted using all auditor characteristics. Standard errors (clustered at the locality or representative vendor level) are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table A.6: **PHASE I – MISCONDUCT: DESCRIPTIVE STATISTICS BY GENDER BASED ON AUDIT TRANSACTIONAL EXERCISES**

		Females		**Males**	
Transaction group	Outcome variable	Mean	SD	Mean	SD
OTC-base	**1**[Misconduct=Yes]	0.62	0.486	0.34	0.477
	Overcharged [GHS]	3.46	1.599	3.71	1.391
OTC-token	**1**[Misconduct=Yes]	0.17	0.383	0.16	0.371
	Overcharged [GHS]	3.25	1.783	3.25	1.949
Falsification	**1**[Misconduct=Yes]	0.08	0.278	0.05	0.235
	Overcharged [GHS]	3.00	1.914	2.12	1.356
Open-account	**1**[Misconduct=Yes]	0.20	0.410	0.22	0.424
	Overcharged [GHS]	3 .00	1.732	2.63	1.120
Overall × Gender	**1**[Misconduct=Yes]	0.28	0.451	0.19	0.395
	Overcharged [GHS]	3.34	1.642	3.31	1.555
Overall	**1**[Misconduct=Yes]	0.23	0.419		
	Overcharged [GHS]	3.32	1.591		

Note: Table shows the descriptive statistics of financial misconduct. These misconduct outcomes are based on the on the pilot experiment transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation or active verification from the side of the customer. The groupings and specific transactions in each transaction group are described in Table A.8. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Overall, the incidence of misconduct is 23% (28% for female vendors; 19% for male vendors) and the average overcharged-amount due to misconduct is GHS3.32 (GHS3.34 for female vendors; GHS3.31 for female vendors).

Table A.7: **PHASE II – MISCONDUCT: DESCRIPTIVE STATISTICS BY GENDER BASED ON AUDIT TRANSACTIONAL EXERCISES**

		Females		**Males**	
Transaction group	Outcome variable	Mean	SD	Mean	SD
OTC-base	**1**[Misconduct=Yes]	0.49	0.501	0.27	0.447
	Overcharged [GHS]	1.86	2.304	2.15	2.036
OTC-token	**1**[Misconduct=Yes]	0.40	0.493	0.27	0.447
	Overcharged [GHS]	2.59	2.068	2.56	2.210
Falsification	**1**[Misconduct=Yes]	0.17	0.382	0.07	0.270
	Overcharged [GHS]	1.96	2.340	1.31	1.108
Open-account	**1**[Misconduct=Yes]	0.35	0.482	0.27	0.449
	Overcharged [GHS]	2.59	1.896	2.17	2.075
Overall × Gender	**1**[Misconduct=Yes]	0.36	0.481	0.22	0.416
	Overcharged [GHS]	2.59	1.896	2.22	2.057
Overall	**1**[Misconduct=Yes]	0.277	0.448		
	Overcharged [GHS]	2.19	2.124		

Note: Table shows the descriptive statistics of financial misconduct. These misconduct outcomes are based on the on the main experiment transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation or active verification from the side of the customer. The groupings and specific transactions in each transaction group are described in Table A.8. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Overall, the incidence of misconduct is 27% (36% for female vendors; 22% for male vendors) and the average overcharged-amount due to misconduct is GHS2.19 (GHS2.59 for female vendors; GHS2.22 for female vendors).

Table A.8: **PHASE I – MISCONDUCT: DESCRIPTIVE STATISTICS BASED ON AUDIT TRANSACTIONAL EXERCISES**

# Transaction type (description)	Outcome variable	Mean	SD	Transaction group	Mean	SD
01 Cash-in GHS50 - to others wallet	**1**[Misconduct=Yes]	0.35	0.480		0.44	0.498
	Overcharged [GHS]	4.65	1.093		3.58	1.498
02 Cash-in GHS160 - to others wallet	**1**[Misconduct=Yes]	0.52	0.502	$\{= \quad OTC - base$		
	Overcharged [GHS]	4.07	0.269			
03 Cash-in GHS1100 - to others wallet	**1**[Misconduct=Yes]	0.48	0.504			
	Overcharged [GHS]	1.85	1.406			
04 Send GHS50 token - to others	**1**[Misconduct=Yes]	0.18	0.390		0.16	0.374
	Overcharged [GHS]	3.68	1.624		3.25	1.850
05 Send GHS1100 token - to others	**1**[Misconduct=Yes]	0.19	0.397			
	Overcharged [GHS]	3.25	1.982	$\{= \quad OTC - token$		
06 Receive GHS50 token - from others	**1**[Misconduct=Yes]	0.20	0.405			
	Overcharged [GHS]	2.71	2.138			
07 Receive GHS1100 token-from others	**1**[Misconduct=Yes]	0.08	0.287			
	Overcharged [GHS]	3.33	2.081			
08 Cash-in GHS50 - to own wallet	**1**[Misconduct=Yes]	0.07	0.259		0.06	0.252
	Overcharged [GHS]	3.20	2.049		2.53	1.641
09 Cash-in GHS160 - to own wallet	**1**[Misconduct=Yes]	0.08	0.274	$\{= \quad Falsification$		
	Overcharged [GHS]	2.00	1.549			
10 Cash-out GHS50 - from own wallet	**1**[Misconduct=Yes]	0.05	0.223			
	Overcharged [GHS]	2.50	1.290			
11 Purchase new SIM card	**1**[Misconduct=Yes]	0.32	0.473		0.21	0.416
	Overcharged [GHS]	2.73	1.099	$\{= \quad Open - account$	2.77	1.352
12 Register new M-Money wallet	**1**[Misconduct=Yes]	0.08	0.280			
	Overcharged [GHS]	3.00	2.645			
Overall	**1**[Misconduct=Yes]	0.23	0.419		0.23	0.419
	Overcharged [GHS]	3.32	1.591		3.32	1.591
Number of transactions		663-1,548			663-1,548	

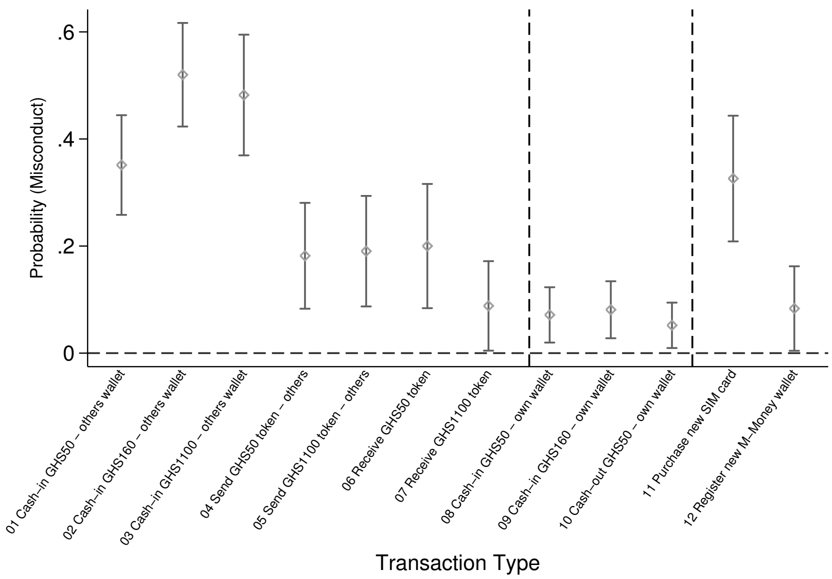
Note: Table reports the specific transactions used for the actual transactional exercises and shows the descriptive statistics of financial misconduct. These misconduct outcomes are based on the the pilot experiment transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation or active verification from the side of the customer. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Overall, the incidence of misconduct is 23% [SD=0.419] and the average overcharged-amount due to misconduct is GHS3.32 [SD=1.591].

Table A.9: **PHASE II – MISCONDUCT: DESCRIPTIVE STATISTICS BASED ON AUDIT TRANSACTIONAL EXERCISES**

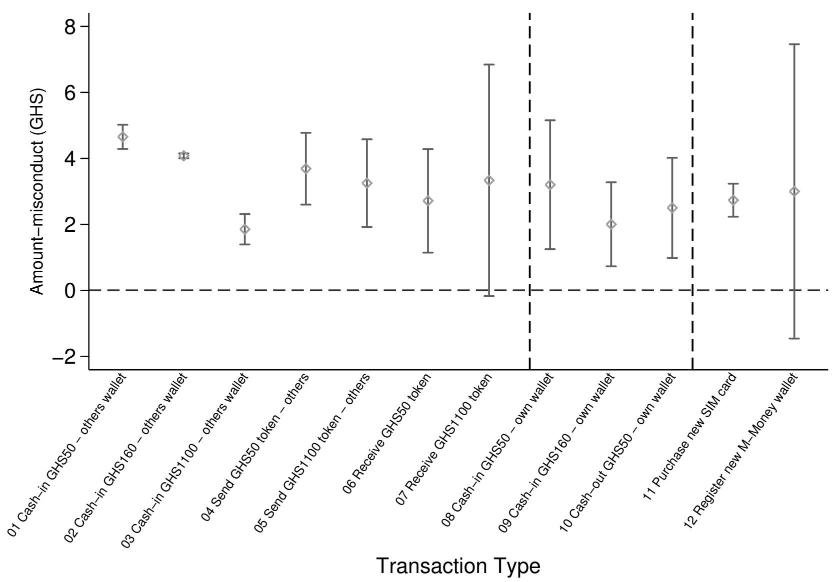
# Transaction type (description)	Outcome variable	Mean	SD	Transaction group	Mean	SD
01 Cash-in GHS50 - to others wallet	**1**[Misconduct=Yes]	0.44	0.498		0.37	0.469
	Overcharged [GHS]	1.64	2.003		2.00	2.192
02 Cash-in GHS160 - to others wallet	**1**[Misconduct=Yes]	0.28	0.453	$\{= \quad OTC - base$		
	Overcharged [GHS]	1.92	1.737			
03 Cash-in GHS1100 - to others wallet	**1**[Misconduct=Yes]	0.38	0.498			
	Overcharged [GHS]	2.53	2.121			
04 Send GHS50 token - to others	**1**[Misconduct=Yes]	0.38	0.490		0.32	0.469
	Overcharged [GHS]	2.39	2.121		2.49	2.133
05 Send GHS1100 token - to others	**1**[Misconduct=Yes]	0.30	0.463			
	Overcharged [GHS]	2.48	2.499			
06 Receive GHS50 token - from others	**1**[Misconduct=Yes]	0.34	0.477	$\{= \quad OTC - token$		
	Overcharged [GHS]	2.88	2.078			
07 Receive GHS1100 token-from others	**1**[Misconduct=Yes]	0.24	0.431			
	Overcharged [GHS]	2.01	1.799			
08 Cash-in GHS50 - to own wallet	**1**[Misconduct=Yes]	0.25	0.440		0.11	0.320
	Overcharged [GHS]	2.25	2.500		1.72	1.984
09 Cash-in GHS160 - to own wallet	**1**[Misconduct=Yes]	0.07	0.260	$\{= \quad Falsification$		
	Overcharged [GHS]	1.00	0.00			
10 Cash-out GHS50 - from own wallet	**1**[Misconduct=Yes]	0.05	0.237			
	Overcharged [GHS]	1.00	0.00			
11 Purchase new SIM card	**1**[Misconduct=Yes]	0.22	0.418		0.31	0.464
	Overcharged [GHS]	1.31	1.126	$\{= \quad Open - account$	2.38	1.981
12 Register new M-Money wallet	**1**[Misconduct=Yes]	0.43	0.498			
	Overcharged [GHS]	3.13	2.115			
Overall	**1**[Misconduct=Yes]	0.277	0.448		0.277	0.448
	Overcharged [GHS]	2.19	2.124		2.19	2.124
Number of transactions		1,181 vs 328			1,181 vs 328	

Note: Table reports the specific transactions used for the actual transactional exercises and shows the descriptive statistics of financial misconduct. These misconduct outcomes are based on the main experiment transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation or active verification from the side of the customer. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Overall, the incidence of misconduct is 27% [SD=0.448] and the average overcharged-amount due to misconduct is GHS2.19 [SD=2.124].

Figure A.3: **PHASE I – MISCONDUCT: DESCRIPTIVE STATISTICS BASED ON AUDIT TRANSACTIONAL EXERCISES**



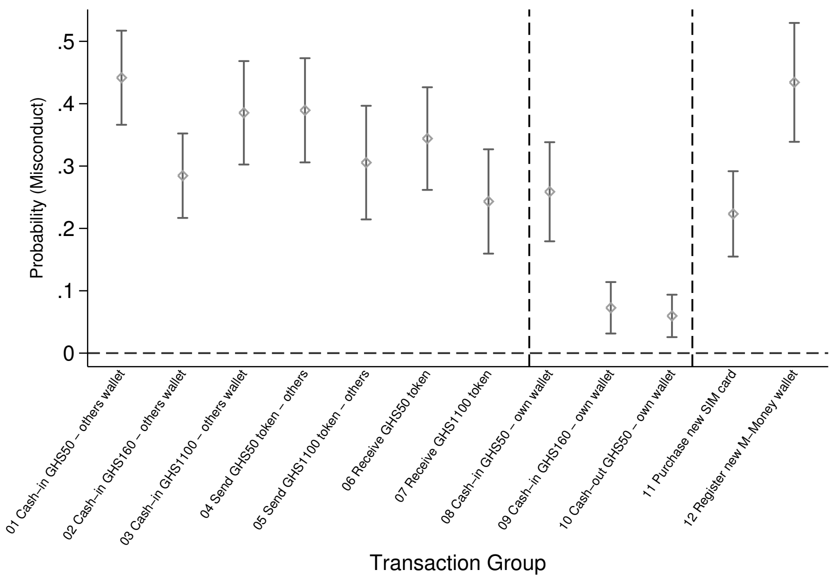
(a) **MISCONDUCT INCIDENCE × TRANSACTION TYPE**



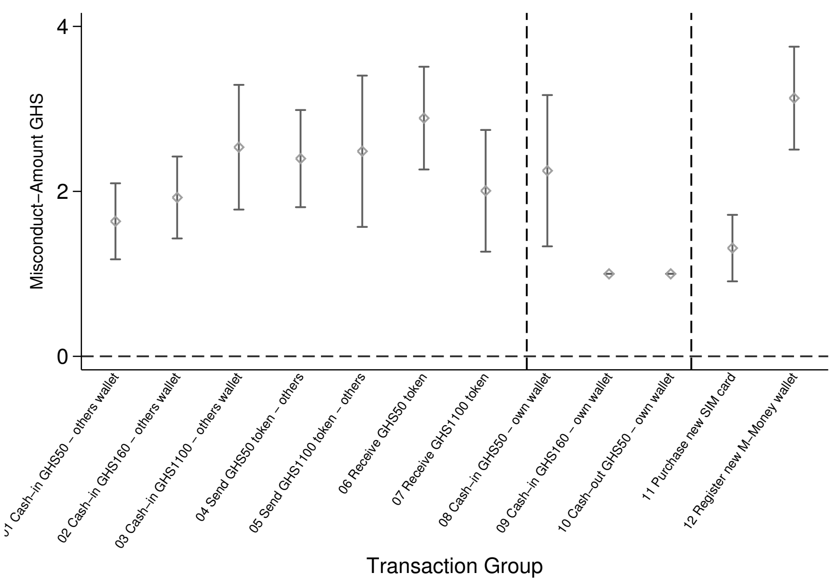
(b) **MISCONDUCT SEVERITY OR AMOUNT × TRANSACTION TYPE**

Note: Figures display the distribution of financial misconduct for the two outcomes (incidence and severity). These misconduct outcomes are based on the pilot experiment transactional exercises. Details of the specific transactions (01-12) are contained in Table A.8. 90% confidence intervals are displayed around the estimates.

Figure A.4: **PHASE II – MISCONDUCT: DESCRIPTIVE STATISTICS BASED ON AUDIT TRANSACTIONAL EXERCISES**



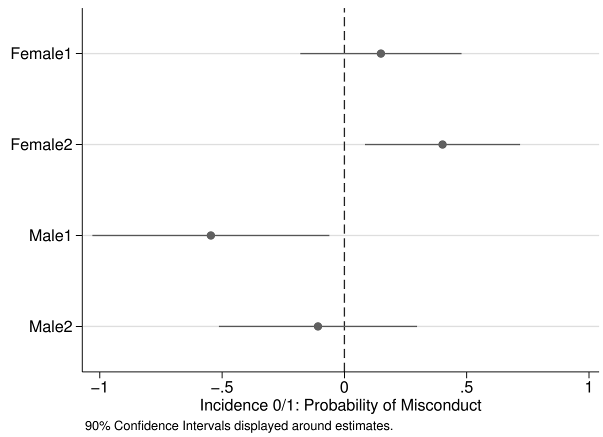
(a) **MISCONDUCT INCIDENCE × TRANSACTION TYPE**



(b) **MISCONDUCT SEVERITY OR AMOUNT × TRANSACTION TYPE**

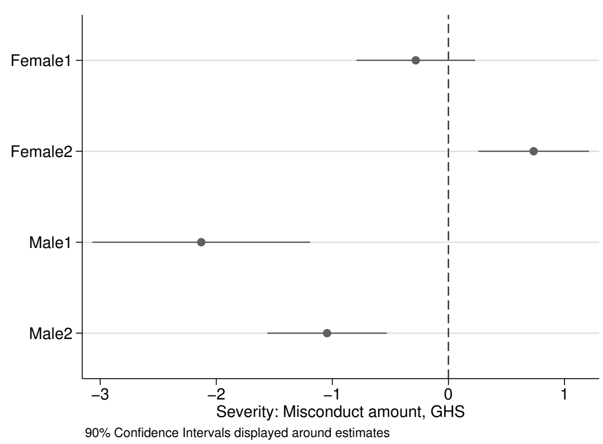
Note: Figures display the distribution of financial misconduct for the two outcomes (incidence and severity). These misconduct outcomes are based on the main experiment transactional exercises. Details of the specific transactions (01-12) are contained in Table A.8. 90% confidence intervals are displayed around the estimates.

Figure A.5: **PHASE I – AUDITOR-SPECIFIC DISTRIBUTION OF MISCONDUCT**



Note: Figure shows auditor-specific plots of misconduct based on a regression of misconduct (Incidence 0/1) against the individual auditor dummies controlling for transaction x date fixed effects. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Severity (Amount-Misconduct) is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. The effects are from gender: systematically, more misconduct is committed against females (female1 and female2) compared to males (male1 and male2).

Figure A.6: **PHASE I – AUDITOR-SPECIFIC DISTRIBUTION OF MISCONDUCT**



Note: Figure shows auditor-specific plots of misconduct based on a regression of misconduct (Severity) against the individual auditor dummies controlling for transaction x date fixed effects. Severity (Amount-Misconduct) is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct at date t. The effects are from gender: systematically, more misconduct is committed against females (female1 and female2) compared to males (male1 and male2).

B Further Results: Mechanisms

I. Main Experiment Transactions, 2021

Table B.1: **Gender and misconduct gap**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: `asinh` (Amount-Misconduct)		
Vendor:	0.138***	0.105**	0.101***	0.189***	0.124*	0.136***
Female (β)	(0.0325)	(0.0442)	(0.0311)	(0.0614)	(0.0670)	(0.0498)
Observations	1,165	1,165	972	1,007	1,007	878
Locality FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.277	0.277	0.277	0.403	0.403	0.403
Number of localities	36	35	35	36	36	35
Number of auditors	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)

Note: Table shows the effects of vendors' gender on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. `asinh` denotes inverse hyperbolic sine. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual locality and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level over the period 3/2021-4/2021. Clustered standard errors (at the local market level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are robust to clustering at either the auditor level or the vendor level.

II. Pilot Experiment Transactions, 2019

Table B.2: **Gender and misconduct gap**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Vendor:	0.0622	0.0790**	0.0851***	0.300**	0.280**	0.295***
Female (β)	(0.0382)	(0.0324)	(0.0311)	(0.129)	(0.121)	(0.113)
Observations	942	942	936	867	867	861
District FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.239	0.239	0.239	0.579	0.579	0.579
Number of districts (localities)	9 (126)	9 (126)	9 (126)	9 (126)	9 (126)	9 (126)
Number of auditors	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)

Note: Table shows the effects of vendors' gender on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual district and transaction × date fixed effects in the possible control set. Observations are at the vendor × transaction × date level over the period 9/2019-10/2019. Clustered standard errors (at the select vendor level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level).

III. Pooled Transactions, 2019 and 2021

Table B.3: **Gender and misconduct gap**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Vendor:	0.108***	0.0898***	0.0725***	0.330***	0.245***	0.229***
Female (β)	(0.0271)	(0.0215)	(0.0227)	(0.100)	(0.0698)	(0.0777)
Observations	2,107	2,107	1,908	1,874	1,874	1,739
District FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.260	0.260	0.260	0.651	0.651	0.651
Number of districts (localities)	10 (162)	10 (161)	10 (161)	10 (162)	10 (162)	10 (161)
Number of auditors	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)

Note: Table shows the effects of vendors' gender on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual district and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level pooled over the periods 9/2019-10/2019 and 3/2021-4/2021. Clustered standard errors (at the vendor level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are robust to clustering at the auditor level.

I. Main Experiment Transactions, 2021

Table B.4: **Gender and misconduct asymmetry – i**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: `asinh` (Amount-Misconduct)		
Customer Assignment: Female (β)	0.0986**	0.100**	0.110**	0.284***	0.229***	0.138
	(0.0403)	(0.0397)	(0.0495)	(0.0617)	(0.0666)	(0.0841)
Observations	1,181	1,181	972	1,007	1,007	878
Locality FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	No	No	No	No	No
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.277	0.277	0.277	0.403	0.403	0.403
Number of localities	36	35	35	36	36	35
Number of auditors	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)

Note: Table shows the impacts of customers' gender assignment on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. `asinh` denotes inverse hyperbolic sine. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, 0-1 indicator for whether auditor is related to the vendor visited or not, and vendor's gender. The double-post LASSO specifications consider all vendor controls, and individual locality and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level over the period 3/2021-4/2021. Clustered standard errors (at the local market level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are robust to clustering at either the auditor level or the vendor level.

II. Pilot Experiment Transactions, 2019

Table B.5: **Gender and misconduct asymmetry – i**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Customer Assignment: Female (β)	0.0265	0.223**	0.223**	-0.156	0.669	0.669*
	(0.0353)	(0.0965)	(0.0909)	(0.121)	(0.407)	(0.382)
Observations	942	942	936	867	867	861
District FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	No	No	No	No	No
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.239	0.239	0.239	0.579	0.579	0.579
Number of districts (localities)	9 (126)	9 (126)	9 (126)	9 (126)	9 (126)	9 (126)
Number of auditors	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)

Note: Table shows the impacts of customers' gender assignment on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, 0-1 indicator for whether auditor is related to the vendor visited or not, and vendor's gender. The double-post LASSO specifications consider all vendor controls, and individual district and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level over the period 9/2019-10/2019. Clustered standard errors (at the select vendor level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level).

III. Pooled Transactions, 2019 and 2021

Table B.6: **Gender and misconduct asymmetry – i**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Customer Assignment: Female (β)	0.0681***	0.0974**	0.101**	0.246**	0.460***	0.453***
	(0.0259)	(0.0378)	(0.0426)	(0.0985)	(0.135)	(0.139)
Observations	2,123	2,123	1,908	1,874	1,874	1,739
District FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.260	0.260	0.260	0.651	0.651	0.651
Number of districts (localities)	10 (162)	10 (161)	10 (161)	10 (162)	10 (162)	10 (161)
Number of auditors	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)

Note: Table shows the impacts of customers' gender assignment on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, 0-1 indicator for whether auditor is related to the vendor visited or not, and vendor's gender. The double-post LASSO specifications consider all vendor controls, and individual district and transaction $\times$ date fixed effects in the possible control set. Observations are at the vendor $\times$ transaction $\times$ date level pooled over the periods 9/2019-10/2019 and 3/2021-4/2021. Clustered standard errors (at the vendor level) are reported in parentheses. *** $p<0.01$ (1% level), ** $p<0.05$ (5% level), * $p<0.1$ (10% level). Standard errors are robust to clustering at the auditor level.

I. Main Experiment Transactions, 2021

Table B.7: **Gender and misconduct asymmetry – II**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: `asinh` (Amount-Misconduct)		
Female vendor:	0.240***	0.174***	0.195***	0.457***	0.322***	0.225***
Female customer Match (β_1)	(0.0469)	(0.0521)	(0.0534)	(0.0692)	(0.0809)	(0.0790)
Female vendor:	0.150***	0.0820	0.127*	0.145**	0.0483	0.0441
Male customer Match (β_2)	(0.0426)	(0.0599)	(0.0741)	(0.0629)	(0.0906)	(0.0985)
Male vendor:	0.113**	0.0870	0.145**	0.250***	0.186*	0.158
Female customer Match (β_3)	(0.0483)	(0.0517)	(0.0636)	(0.0784)	(0.0998)	(0.117)
Observations	1,181	1,181	972	1,007	1,007	878
Locality FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	No	No	No	No	No
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.277	0.277	0.277	0.403	0.403	0.403
Number of localities	36	35	35	36	36	35
Number of auditors	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)	40 (20M; 20F)
p-value (test: $\beta_1 = \beta_2$)	0.070	0.090	0.296	0.000	0.002	0.073
p-value (test: $\beta_1 = \beta_3$)	0.004	0.031	0.244	0.011	0.095	0.449

Note: Table shows the impacts of random gender matches between customers and vendors on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. `asinh` denotes inverse hyperbolic sine. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual locality and transaction × date fixed effects in the possible control set. Observations are at the vendor × transaction × date level over the period 3/2021-4/2021. Clustered standard errors (at the local market level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are robust to clustering at either the auditor level or the vendor level.

II. Pilot Experiment Transactions, 2019

Table B.8: **Gender and misconduct asymmetry – II**

	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Female vendor:	0.0957	0.279***	0.279***	0.148	0.843**	0.843**
Female customer Match (β_1)	(0.0647)	(0.0965)	(0.0908)	(0.207)	(0.410)	(0.383)
Female vendor:	0.0928**	0.133***	0.133***	0.321**	0.508***	0.508***
Male customer Match (β_2)	(0.0448)	(0.0496)	(0.0466)	(0.157)	(0.175)	(0.164)
Male vendor:	0.0562	0.246***	0.246***	-0.0819	0.777**	0.777**
Female customer Match (β_3)	(0.0399)	(0.0912)	(0.0858)	(0.136)	(0.377)	(0.353)
Observations	942	942	936	867	867	861
District FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	No	No	No	No	No
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.239	0.239	0.239	0.579	0.579	0.579
Number of districts (localities)	9 (126)	9 (126)	9 (126)	9 (126)	9 (126)	9 (126)
Number of auditors	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)	4 (2M; 2F)
p-value (test: $\beta_1 = \beta_2$)	0.967	0.150	0.124	0.450	0.422	0.389
p-value (test: $\beta_1 = \beta_3$)	0.555	0.454	0.424	0.287	0.652	0.629

Note: Table shows the impacts of random gender matches between customers and vendors on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual district and transaction × date fixed effects in the possible control set. Observations are at the vendor × transaction × date level over the period 9/2019-10/2019. Clustered standard errors (at the select vendor level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level).

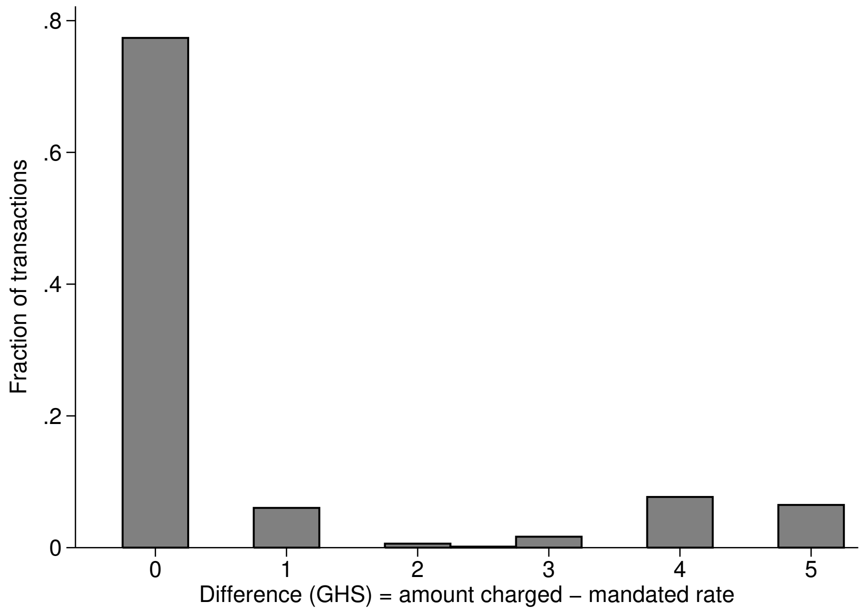
III. Pooled Transactions, 2019 and 2021

Table B.9: **Gender and misconduct asymmetry – II**

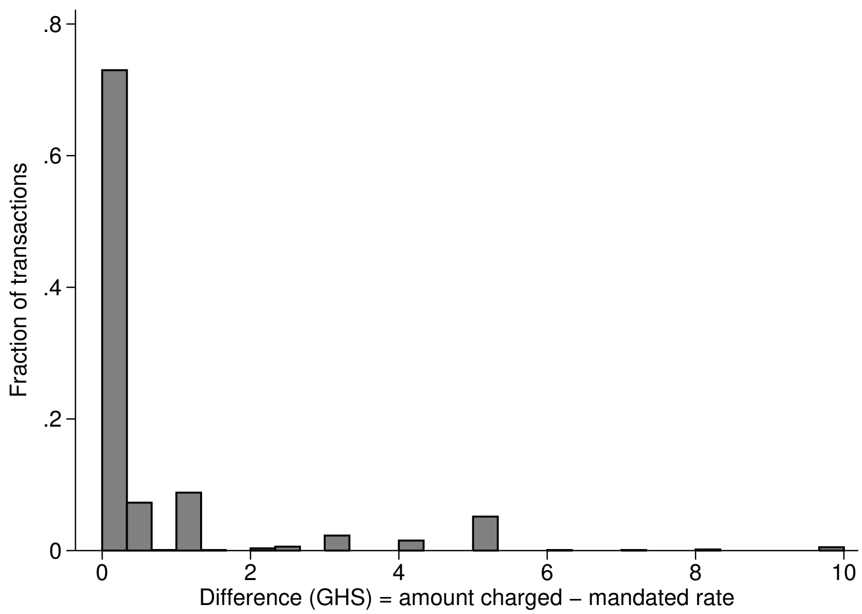
	(1)	(2)	(3)	(4)	(5)	(6)
	Incidence: Misconduct 0-1			Severity: Amount-Misconduct, GHS		
Female vendor:	0.192***	0.185***	0.175***	0.617***	0.674***	0.613***
Female customer Match (β_1)	(0.0394)	(0.0434)	(0.0459)	(0.152)	(0.157)	(0.160)
Female vendor:	0.123***	0.124***	0.126***	0.287***	0.287***	0.285***
Male customer Match (β_2)	(0.0318)	(0.0328)	(0.0340)	(0.0978)	(0.0977)	(0.0975)
Male vendor:	0.0863***	0.109***	0.138***	0.217*	0.494***	0.524***
Female customer Match (β_3)	(0.0296)	(0.0396)	(0.0467)	(0.113)	(0.159)	(0.164)
Observations	2,123	2,123	1,908	1,874	1,874	1,739
District FE	No	Yes	Yes	No	Yes	Yes
Transaction x Date FE	No	Yes	Yes	No	Yes	Yes
Auditor FE	No	Yes	Yes	No	Yes	Yes
Controls	None	None	Post-Double LASSO	None	None	Post-Double LASSO
Mean of dependent variable	0.260	0.260	0.260	0.651	0.651	0.651
Number of districts (localities)	10 (162)	10 (161)	10 (161)	10 (162)	10 (162)	10 (161)
Number of auditors	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)	44 (22M; 22F)
p-value (test: $\beta_1 = \beta_2$)	0.115	0.198	0.326	0.044	0.015	0.040
p-value (test: $\beta_1 = \beta_3$)	0.011	0.026	0.283	0.019	0.185	0.513

Note: Table shows the impacts of random gender matches between customers and vendors on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: age, 0-1 indicator for whether married or not, 0-1 indicator for whether belongs to akan ethnic group or not, 0-1 indicator for whether self employed or not, experience in business, business size, 0-1 indicator for whether official tariff was posted or not, indicator for whether involved in other non-mobile money business, wait time for transaction, and 0-1 indicator for whether auditor is related to the vendor visited or not. The double-post LASSO specifications consider all vendor controls, and individual district and transaction × date fixed effects in the possible control set. Observations are at the vendor × transaction × date level pooled over the periods 9/2019-10/2019 and 3/2021-4/2021. Clustered standard errors (at the vendor level) are reported in parentheses. *** p<0.01 (1% level), ** p<0.05 (5% level), * p<0.1 (10% level). Standard errors are robust to clustering at the auditor level.

Figure B.1: **DISTRIBUTION OF DIFFERENCE BETWEEN OBSERVED CHARGES AND MANDATED RATES**



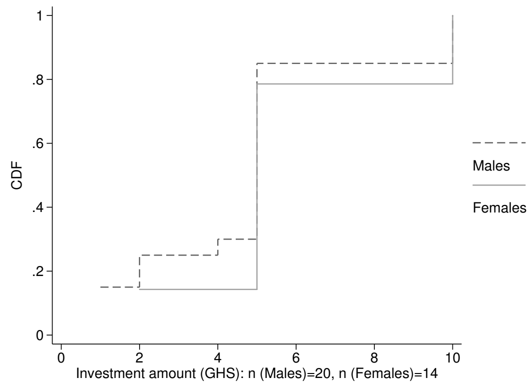
(a) **PHASE I–PILOT EXPERIMENT: OBSERVED AND MANDATED RATES**



(b) **PHASE II–MAIN EXPERIMENT: OBSERVED AND MANDATED RATES**

Note: Figure shows the distribution of actual transactional charges relative to the mandated rates separately for the pilot and main experiments. This measures the likelihood of under-charging (if the difference is negative), correct-charging (if the difference equal to 0), and over-charging (if the difference is positive). Calculations are based on transaction data from the field trials.

Figure B.2: **RISKY INVESTMENTS BY GENDER**



Notes: Figure shows the distribution of investment choices or amounts (GHS) by vendors (females *versus* males) to an investment game meant to elicit their risk-attitudes. These investment choices provide an estimate of risk aversion for each vendor, whereby the higher the investment amount the less risk averse is the vendor (see e.g., Gneezy and Potters 1997). There is limited graphical evidence that the two distributions are significantly different (also consistent with results from a formal two-sample Kolmogorov–Smirnov test of distributional equality). A simple regression of the investments on an indicator for whether the vendor is a female or not provides a p-value= 0.183, suggesting that the two distributions are not significantly different from each other.

C Elicitation: Beliefs about Gender

Market Beliefs about Misconduct and Gender

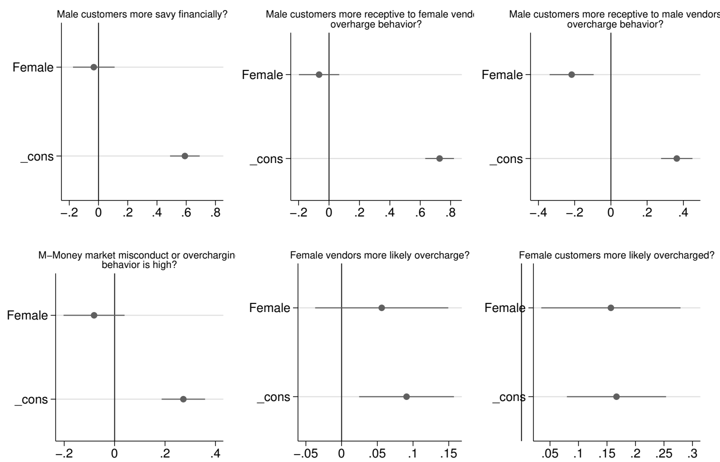
Between April-May 2020, we conducted a wave of phone survey (due to COVID-19 disruptions) to elicit market beliefs, capturing perceptions about various aspects of misconduct on M-Money. For each of the 6 statements below, market participants were asked to indicate their belief (i.e., `Agree`/ `Disagree`). The respondents consist of a representative sample of 32 local markets: 32 vendors and 182 nearby customers (drawn from our Baseline I market census). The statements were designed to reflect the gender-differentiated market facts obtained from the pilot field trials.

For each of the statements, subjects were jointly asked to guess the percentage of others (all vendors and customers in their locality) that will `Agree` to the statement (i.e., beliefs about others beliefs). To incentivize their reports, among all respondents in a locality, "the respondent" with the closest guess (to the locality-level estimate) immediately received 10GHS after all respondents have answered either in-cash through their M-Money or in-kind through a phone calling-credit. All respondents were informed of this payoff before they answered. Table C.1 outlines the specific statements.

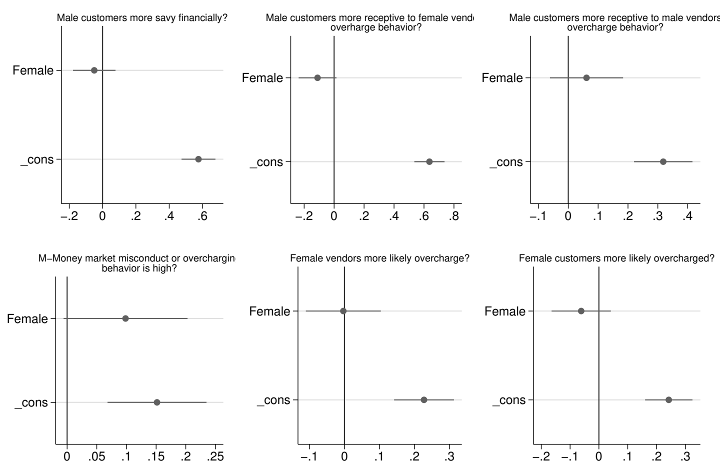
Table C.1: **BELIEF STATEMENTS ABOUT GENDER AND MISCONDUCT**

No.	Statement
01a	In [my] view, M-Money Male-customers are more sophisticated or "savvy" financially ... than Female- customers? `1=Agree, 2=Disagree`
01b	What's [your] estimate of the % of others (all vendors and customers in this locality) that ... will Agree with [01a]? -----%
02a	In [my] view, M-Money Male-customers are more receptive than Female- customers to being ... "over- charged above mandated charges" by Female- vendors? `1=Agree, 2=Disagree`
02b	What's [your] estimate of the % of others (all vendors and customers in this locality) that will Agree with [02a]? -----%
03a	In [my] view, M-Money Male-customers are more receptive than Female-customers to being ... "over- charged above mandated charges" by Male-vendors? `1=Agree, 2=Disagree`
03b	What's [your] estimate of the % of others (all vendors and customers in this locality) that will Agree with [03a]? -----%
04a	In [my] view, general misconduct or overcharging customers' transactions at M-Money ... vendor points is high? `1=Agree, 2=Disagree`
04b	What's [your] estimate of the % of others (all vendors and customers in this locality) that will Agree with [04a]? -----%
05a	In [my] view, Female- vendors are more likely than Male- vendors to "overcharge" ... customers at M-Money vendor points? `1=Agree, 2=Disagree`
05b	What's [your] estimate of the % of others (all vendors and customers in this locality) that will Agree with [05a]? -----%
06a	In [my] view, Female- customers are more likely to be "overcharged" at M-Money ... vendor points? `1=Agree, 2=Disagree`
06b	What's [your] estimate of the % of others (all vendors and customers in this locality) that will Agree with [06a]? -----%

Figure C.1: **Differences in beliefs about misconduct dimensions**



(a) **vendors**



(b) **customers**

Note: Figure shows the differences in beliefs across 6 selected statements about misconduct from a linear probability model (by gender and market participant-type). For each of the statements, market participants were asked to indicate their belief (i.e., `Agree`/ `Disagree`). Details about the statements are contained in Table C.1. The dependent variable is a dummy variable indicating whether or not the respondent `Agree` with the statement. Female is a 0-1 indicator for whether respondent is a female or not. Coefficients are in percentage points. Observations are at the market individual level (32 vendors; 182 customers). 90% confidence intervals are displayed around the estimates for statistical significance.

D Auditors' Training

INSTRUCTIONS:
VENDOR-BASED APPROVED TRANSACTION TARIFFS

- Welcome: You have been "assigned" to vendor shops, where you will make specific Mobile Money transactions.
- You will be required to use the same language while transacting at vendor shops (details below).
- Our focus will be vendor- or merchant-based Mobile Money transactions.
- Throughout, we pay fees whenever we are sending money at the vendor to guarantee the receiver receives XGHS-amount.
- Most at times picking up money from the vendor should be free (details below).
- Here are the approved rates that we will be working or transacting with at vendors' premises (Let's memorize them. You will be given copies, so you can refer these rates any time you are in doubt):

KEY: TRANSACTIONAL CODES
OVER-THE-COUNTER, OTC

- T1: Put GHS50 on someone's (XX/Yourselves) M-Money wallet {GHS50 => PAY GHS0.5}
- T2: Put GHS160 on someone's (XX/Yourselves) M-Money wallet {GHS160 => PAY GHS1.6}
- T3: Put GHS1100 on someone's (XX/Yourselves) M-Money wallet {GHS1100 => PAY GHS10}

TOKEN

- T4: Send a Token of GHS50 to someone (XX/Yourselves) {GHS50 => PAY GHS2.5}
- T5: Send a Token of GHS1100 to someone (XX/Yourselves) {GHS1100 => PAY GHS55}
- T6: Receive a Token of GHS50 from someone (XX/Yourselves) **{GHS50 => FREE}
- T7: Receive a Token of GHS1100 from someone (XX/Yourselves) **{GHS1100 => FREE}

FALSIFY [INSTANT VERIFIABILITY PROVIDED BY PROVIDER]

- T8: Put or Cash-in GHS50 on your own M-Money wallet {GHC50 => FREE}
- T9: Put or Cash-in GHS110 on your own M-Money wallet {GHS110 => FREE}
- T10: Take or Cash-out GHS50 from your own M-Money wallet {GHS50 => FREE}

ACCOUNT OPENING

- T11: Buy a new SIM card {SIM (or ATTEMPT it) => PAY GHS2}
- T12: Then use T11 to register for Mobile Money Account {REGISTER (or ATTEMPT it) => FREE; initial deposit of GHS5 minimum required but this GHS5 must be on your account, merchant should not take it, verify}.

TRANSACTION APPROACH

**DURING VISIT (Very simple language, no deviations allowed): `Good morning /afternoon /evening. I want to make a M-Money transaction [USE CODES: T1...T12].`

- `Present necessary details: phone number, and sender or recipient details`
- `Thank you for your service`

**AFTER VISIT: Immediately complete the questionnaire (see Table D.1) right after the transaction using your Tablets.

ADDITIONAL NOTES

- [1] The order of transactions to make at vendor points will always be determined (randomly) by the CAPI data entry software on your Tablets (you don't choose it). CAPI will also display the various tariffs in case you are in doubt.
- [2] Please leave spaces blank *if* a specific transaction-type is not feasible (the software will randomly switch to another transaction-type).
- [3] Practicing: let's take turns to practice repeatedly the transaction approach, using yourselves as vendors and other nearby M-Money vendors. Your supervisors will be monitoring... Any questions or clarifications? Let's discuss.

Table D.1: **QUESTIONNAIRE: AUDITOR'S UNIQUE ID. . .**

Q0	Q1a	Q1b	Q1c	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
No.	VISIT DATE			Locality	"Rep"	TRANSACTION	Transaction	How much	Transaction	Appx wait time	Related to	How are you related to	Vendor's Gender?	Vendor involved
	MM	DD	TIME	code?	Vendor	TYPE? USE	OVERCHARGED?	difference?	successful?	transaction	Vendor just visited?	Vendor? 1=RELATIVE;	1=MALE	in non-Mobile Money
					code?	CODES: T1...T12	1=YES; 2=NO=>Q7	GHS	1=YES 2=NO	took? MINS	1=YES; 2=NO => Q11	2=FRIEND; 3=OTHER	2=FEMALE	businesses? 1=YES 2=NO
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
.														
.														
.														
.														
.														
.														

. . . . . .