GENDER AND FINANCIAL MISCONDUCT: A FIELD EXPERIMENT ON MOBILE MONEY

Francis Annan*

Georgia State University

MARCH 2021

Abstract

We construct a census of the market for mobile banking in village Ghana and estimate that 1 out of every 5 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 40% more likely to commit such misconduct relative to male vendors. Misconduct is asymmetric: female customers are 89-96% relatively more likely to suffer misconduct, and while female vendors discriminate against customers of their gender, male vendors favor their gender. Differences in empowerment and beliefs about gender 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`. 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, and seminar participants for their helpful comments. I thank UChicago Advances with Field Experiments and UWestern Ontario 3rd Toronto Fintech conference participants for comments on an earlier draft of this paper. Excellent support from the field coordinator, Samuel Adotevi (of Ghana Statistical Service), is acknowledged. 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.

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 (see, e.g., 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. 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 the first study that examines the nature of misconduct in markets and quantifies its gender impacts, drawing on the local market for mobile money [M-Money] in Ghanaian villages. Our general focus on gender is motivated by existing research showing gender differences in market-driven primitives such as differences in risk attitudes, financial investments and sophistication, competition, social (other-regarding) preferences, beliefs, etc (Charness and Gneezy 2012; Sunden and Surette 1998, Bannier and Neubert 2016; Reuben, Sapienza and Zingales 2015; Croson and Gneezy 2009; Glover, Pallais and Pariente 2017, Bordalo et al. 2019, respectively). Potential gender differences in misconduct punishment may also exist (Egan, Matvos and Seru 2019), with impacts on financial and labor market outcomes.

M-Money is an important financial innovation in developing economies and a well-celebrated example of FinTech (Bharadwaj, Jack and Suri 2019; Goldstein, Jiang and Karolyi 2019), with much promise for financial development and inclusion. It provides financial services and transactions which are delivered on digital mobile networks, and comprises market vendors, who are small business outlets that provide cash-in and cash-out services, 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 financial misconduct based on two appealing features: compared to traditional banking, [i] it is less regulated, and [ii] 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 convenient access and about 90% of households, their close family and friend networks have registered for a M-Money account. Transactional charges and practices are "officially" set or defined by the regulator and providers that the vendors work for. We define misconduct to entail all transactions at the vendor banking point that are indicative of fraud or wrongdoing. Here, this captures over-charging and/or faking transactions with reference to the regulator and provider-approved charges and practices.

In this environment, financial misconduct is especially an undesirable outcome because it can be discriminatory and imposes significant financial burden on households. For example, we estimated that the average overcharge due to misconduct reflects about 82% of mean official charges, with disproportionate burden on female customers. The potential efficiency costs of misconduct can be quite severe. The over-charged funds may go into unproductive activities; in practice, this can occur if female-vendors transfer their over-charged funds or parts of it (from female-customers) to their husbands and other members (see, Bernhardt et al. 2019 for relevant evidence in Ghana and Sri Lanka) who may spend on less productive activities (Duflo 2003). 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. Thus, understanding gender differences in misconduct speaks to important issues about discrimination and inefficiency, with implications for policy.

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 137 poor and low-income communities in Eastern Ghana. We deployed trained field officers to visit each of these localities to list all vendors and 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 study the prevalence and nature of misconduct, we go beyond the typical audit methodology by recruiting experimental customers in our study area to act as auditors. We give them cash to make actual transactions on M-Money. 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 (see, Kessler, Low and Sullivan 2019). There is descriptive evidence of significant amount of misconduct: the overall incidence of misconduct is 23% and the average overcharged-amount due to misconduct reflects about 82% of mean official charges. Misconduct is substantially higher for female vendors (28%) compared to their counterpart male vendors (19%).

We take advantage of our set up to manipulate the market match between vendors and customers, which allows for clean measurement of misconduct and identification of gender effects. Our identification strategy exploits exogenous variation created by the random matches between vendors and customers. We verify the validity of our design by re-matching vendors and customers, whereby the initial gender assigned was reversed in a second wave of transactional exercise, and showing that misconduct and estimated effects are prevalent in the space of transactions that are vulnerable to misconduct.

Our baseline empirical model evaluates gender differences in misconduct for vendors, and the relative differences and effects from the mismatch in gender between vendors and customers, while controlling for market district fixed effect, and transaction $\times$ date fixed effect. These fixed effects allow us to compare male and female vendors who do business in the market district, the same transaction type and at the same transaction date, flexibly accounting for any unobservable differences based on location, transaction or market cycles. We find economically and statistically significant effects on both.

There is strong evidence of a "gender misconduct gap": based on randomized matches between vendors and customers, we find that female vendors are 9 percentage points (pp) (equivalently +40%) more likely to commit misconduct relative to male vendors. The nature of financial misconduct is asymmetric: female customers are 89-96% more likely to suffer misconduct relative to similar customers who are males. Relative to a male vendor-male customer match, female vendors are 28 pp more likely to cheat female customers but 13 pp more likely to cheat male customers. In contrast, male vendors are 25 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 male customers. These effects are robust to several alternative model specifications, the influence of customers' gender, the use of post-double-selection LASSO for estimation, and falsification tests.

What explains the gender misconduct gap, market discrimination, within-gender discrimination versus within-gender favoritism? Motivated by existing theoretical and applied research (Charness and Gneezy 2012; Sunden and Surette 1998, Bannier and Neubert 2016; Reuben, Sapienza and Zingales 2015; Croson and Gneezy 2009; Glover, Pallais and Pariente 2017, Bordalo et al. 2019), we investigate several competing hypotheses: differences in risk taking, market concentration effects, market costs of misconduct, effects of price-transparency, empowerment or social distance, and differences in market beliefs.

For risk taking: we re-visited the vendors and measured their risk attitudes using an investment-based measure of risk aversion (Gneezy and Potters 1997). We find no significant differences by gender, suggesting the limited influence of risk attitudes. For market concentration: we draw on sales data from our baseline market census to construct a Herfindahl-Hirschman index, which also shows no meaningful gender differences in competition (as defined by the Herfindahl-Hirschman index). Next, we formulate a simple model that captures relevant features of both price transparency and market monitoring effects of misconduct to illustrate formally how these two effects could act to affect the incidence of misconduct by gender. We test implications from this model, and reject both market costs of misconduct and price transparency as likely mechanisms.

For female empowerment: our evidence provides the most support for it. We use data on empowerment of women from the most recent Demographic Health Survey (DHS) to construct two common indices of female empowerment (DHS 2014), reflecting the number of decisions that women participate in alone or jointly and reasons for which a husband is justified to beat his wife. This allows us to examine the influence of gender empowerment at the market district level, and we find that low women empowerment drives our estimated effects. We argue this evidence is consistent with the theory that preexisting low female empowerment incentivizes excessive profit maximization motives for female vendors. This creates incentives for more misconduct of the female vendors, which is committed more on female customers than on male customers, who are presumably more empowered than the female vendors. Male vendors who are also more empowered cheat the less empowered female customers. We also personal income data of vendors as an objective proxy for status and women empowerment provide micro-evidence which re-affirms differences in empowerment as a relevant mechanism.

In additional surveys and tests, we emphasize differences in beliefs or perceptions about gender and misconduct between vendors and customers as a relevant parallel mechanism. Specifically, we find evidence that the market views male customers as more sophisticated, and customers' underperceive the level of misconduct by female vendors. Such beliefs are consistent with the estimated gender differences in misconduct, e.g., with why female customers are cheated more and why female vendors act opportunistically and thus commit more misconduct. Finally, we rule out alternative explanations such as gender differences in social cost of misconduct, spatial clustering of female vendors and peer influences based on a plethora of tests. Our findings raise important issues at the intersection of economics and culture, and indicate that preexisting social distance (particularly unequal gender empowerment and gendered beliefs) can incentivize undesirable market outcomes and may be an important source of local financial market frictions.

II Connections to the Literature

We add to several distinct literatures, with implications for policy.

II.1 Financial Misconduct: There are studies documenting higher incidence of misconduct for males in the financial industry and the implied discriminatory responses to punishments (Egan, Matvos and Seru 2019). Our evidence is rather the opposite, showing higher misconduct for females. Egan, Matvos and Seru (2019) focuses on the labor market response of gender differences in misconduct, while we focus on the nature of misconduct documenting asymmetries within-gender which we rationalize with differences in social distance and beliefs about gender. Other studies have explored variation in misconduct across space (termed "financial misconduct puzzle"; see Parsons, Sulaeman and Titman 2018) and in intensity (Karpoff and Lou 2010; Dimmock, Gerken and Graham 2018). These papers have emphasized the importance of socio-local norms (as in Sah 1991; Glaeser, Sacerdote and Scheinkman 1996), particularly peer effects. We complement this work by studying misconduct differences across *gender lines*, and we test and emphasize a different mechanism "social distance" which is also related to local norms. Together, this body of work shows that social-local norms, which can take different forms, play an important role in shaping financial misconduct.

Finally, despite the promise of FinTech (e.g., in poverty reduction, risk sharing, resilience and personal finance, entrepreneurship impacts; see, Bharadwaj, Jack and Suri (2019)), we are arguably the first to document the nature of misconduct using manipulated assignments of market participants from an emerging market setting.

II.2 Market Discrimination: Gender differences in finance and labor outcomes typically include discrimination in wages and hiring. Bertrand and Duflo (2017) provides a review, summarizing differential treatment of race and discrimination in labor markets. Recent studies in India (Banerjee et al. 2009) report minimal experimental evidence of discrimination based on caste, and in Nigeria (Archibong et al. 2019) provide descriptive evidence of discrimination based on gender and ethnicity. There is also evidence showing gender punishment gaps from misconduct or wrongdoing. We complement this literature in two ways. First, our evidence that for female-customers "i.e., the marginalized", the market for M-Money is an uneven playing field reaffirms previous work. Second, the vast available evidence so far suggests that discrimination runs "across groups", and not within-group (Egan, Matvos and Seru 2019; Abbink and Harris 2019, etc). 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.

II.3 Policy Aspects: 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, such as social distance.

II.4 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 23% on incidence and 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). Similarly, our result on within-group favoritism also points to a specific source of corruption where public officials could abuse their power in order to distribute positions or resources to their "own groups" at the expense of the public at large (Abbink and Harris 2019; Fatton 1990, Englebert 2000; Kaufmann et al. 2006).

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

III Setting and Research Design

III.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 [2006; 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 banking points, but to characterize its nature and

[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&

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. Table B.1 shows the timeline of the study.

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.

III.2 Market Census, and Data Set

Market Census: Between February-March 2019, we carried out a unique census of the market for M-Money in Eastern Ghana, spanning 9 districts. Figure B.2 shows our nine experimental districts located in (the southeastern belt of) Eastern Ghana. 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 mobile banking sites, and our initial pilot works (in February 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 banking points, and other surrounding households. We focus on nearby households in order to maximize our chances of studying households that might make transactions with the select vendors.

To focus on low-income environments and to ensure the presence of *at least* a M-Money center 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, which we shall refer to as "local markets". The GPS-recorded polygons of all the selected local markets are shown in Figure B.1. In practice, we find that 130 out of these 137 localities had one or more M-Money center(s) after we undertook the baseline market census (implying a 95% success rate), whereby: trained field officers were deployed to visit each of the selected localities to list all vendors and all nearby customers who are within 5 houses radius around a given vendor.

Thus, a local market is defined to reflect the pair: vendor by the set of all nearby customers.

Market-Wide Stylized 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.[2]

Table 1 shows the summary statistics for the market. To facilitate comparisons between both sides of the market, the relevant statistics for vendors and customers are displayed next to it each other. Female vendorship is 39%, 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, half 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 B.2). 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.[3]

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]) of vendors operate as a joint venture, bundling this with other services.[4] The average daily sales per vendor is about

[2]Detail summaries and other patterns about the market are available, upon request.

[3]Poverty estimates, formally: since our study focuses on mobile banking 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.

[4]We identified joint venture services like: groceries and provisions, local medicine, multi TV installation, registration of SIM

GHS2,260 [US$442] (not statistically different between female and male vendors, see Table B.2). 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 mobile banking site, while this distance is about 383 meters for post-offices. Finally, in Table B.2, we break down the data for vendors by gender, illustrating that female vendors compare well with male vendors on several relevant variables in the market census.

III.3 Research Design

For our purposes, we employ an innovative audit study where auditors (experimental customers; details below) 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 for M-Money, totaling 12: sending versus receiving transactions.

The first feature allows us to credibly study gender differentiated effects, gender discrimination and favoritism, while the second helps to circumvent potential errors 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

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.

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 are demographically similar (e.g., via marital status, and age) to the market's customer distribution, and can speak and act similarly as traditional customers will typically act.

Balance and Validity of Design: Each of the 130 localities has one or more vendor(s), with a range of 1-13 vendors. To maximize statistical power, we "randomly" select one vendor per locality for our field transactional exercises, which we shall refer to as "representative vendors". A combination of these representative vendors and their nearby customers thus generates the "representative markets". To what extent are the random samples of vendors or markets representative of the entire market population? We base our transactional exercises on the representative vendors. Sample representativeness and identification requires that being a representative vendor (i.e., the assignment of which vendor or market is representative) is independent of any relevant market-level statistics. To test that these samples are comparable to the market population, we run the following regression

$$y_i = \alpha + \beta S_i + \epsilon_i$$

on the baseline census data, where $S_i = 1$ if vendor or market i is selected to be a representative 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 B.3 and B.4 report the results, where we find no difference across markets selected and those not-selected to be representative. We compare the distribution of representative vendors by gender in our analysis of mechanisms.

Auditors and Random Assignments: We utilize a total of four auditors or experimental customers (2 males and 2 females) for the market transactions and random assignment to vendors. We use few auditors to simplify the assignments to vendors, minimize cross-auditor heterogeneity while focussing on the gender (mis)matches with vendors.[5] A stratified random assignment was used: 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 ensures equal random assignments to opposite and same gender.

To test whether the randomization was successful, we run regressions (similar to the sample-population comparisons) that compare auditor characteristics based on their assignments to the 130 representative vendors. Table B.5 reports the results, and shows strong evidence of covariate balance and thus successful randomization. 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.

Auditors and Transactional Exercises: The auditors were chosen from our research partner's pool of field officers, reside in our study area and compare well demographically to the population of customers, as noted, 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

[5]There is a tradeoff: with few auditors (2 men and 2 women), it becomes difficult to make inference by consumer or auditor gender; the males and females could happen to differ on some other important trait besides gender. If we plot the distribution of misconduct for each of the four 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 C.5 and C.6). To address this formally, we re-ran the experiment in March 2021 using a large number of auditors (20 men and 20 women). These were assigned to 163 vendors across 36 localities. We find similar effects across both audit designs.

[6]In practice, a very large share of market vendor transactions are conducted with customers who have no family and/or close relations. In Annan (2020), 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.

set of transactions (see details in Appendix D). They 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 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 ensured quality of the transactional exercises in three ways. First, we use research supervisors, who occasionally visit the transaction centers as potential customers waiting in line to transact while the assigned auditors are transacting at the vendor shops. The supervisors observe the auditors' actions and verify their transaction data. Second, we set up a computer-adaptive data collection platform, which allowed us to track and verify the data in real time. Right after every visit, auditors complete a brief questionnaire about the transaction (see, Table E.1 in Appendix D) and synchronize the data to our platform for immediate access. Third, we piloted the proposed audit approach in February 2017, which yielded similar patterns of misconduct (as noted in the Market Census section).

III.4 Misconduct: Measurement and Descriptive Evidence

Our main field transactional exercises cover all the 130 representative vendors, and were carried out between September-October 2019. Multiple transactions are performed at each banking site 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. About 6 successful trips were made per auditor per day to their assigned vendors.

Measuring Misconduct: In our market setting, misconduct of vendors 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 (and as in Egan, Matvos and Seru 2019), we simply define misconduct to entail transactions that are indicative of fraud or wrongdoing i.e., over-charging and/or faking transactions but with reference to the regulator and provider-approved charges and practices. A major advantage of our framework is that we are able to measure misconduct at granular levels using the transactional exercises: across different types of transactions, the specific incidence of it (extensive margin) and amount overcharged as a result of the misconduct (intensive margin).

Table B.7 reports the descriptive statistics of vendors' misconduct overall, and across different transactional classes (the full distributions are provided in Figures B.3 and 1, for additional reference). 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 Table B.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]). In 4 out of the 5 transactional groups, the evidence is similar: female vendors committed more misconduct (except for account opening, where misconduct is 20% for females but 22% for males). Turning to the misconduct outcome on severity, there are (almost) similar patterns: the average overcharged amount due to misconduct is slightly higher for female vendors (GHS3.35 for females; GHS3.31 for males). Female vendors overcharged more on average in 4 out of the 5 groups, except for the OTC-base transactional group where the average amount overcharged is GHS3.46 for females but GHS3.71 for males.

[7]All shown and described in Table B.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 types of transactions or groups. 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). We explore this as a falsification check in the formal analysis of gender gaps in misconduct. Second, misconduct is potentially "costly" to consumers or households. The average false charges due to misconduct reflect about 82% of mean official charges, which may impose additional financial burden on households. 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, eg., DeLiema et al. 2018). In our baseline 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`. 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 1, 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%, 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.

IV Empirical Strategy

IV.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. Within-gender favoritism and across-gender discrimination also create different incentives to influence the misconduct of vendors.[8]

IV.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

[8] By randomly assigning customers, 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. It does not address potential differences in shop attributes by vendors' gender. We explore such differences in vendors gender and shop attributes (i.e., representative vendors who participated in the audit exercises) as potential mechanisms underlying our results.

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, transaction type, and transaction date, respectively. 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. 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 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

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

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. In alternative models, 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.

All standard errors are clustered at the vendor level to account for correlations of transactions within vendor (Cameron and Miller 2015). We will sometimes discuss effects that contain useful economic information (i.e., looking at effect sign and effect size)–whether statistically significant or not (Abadie 2020).

V Results

V.1 Gender Differences in Misconduct

Table 2 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 [Table 2]. 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 [Table 2], 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 9 pp. With an overall misconduct of 23%, the estimated difference corresponds to $\frac{0.09}{0.23} \times 100 = +39\%$ higher misconduct incidence for the female vendors. For the intensity outcome, the effect is similar and corresponds to +40%.

Conditional on the market district, and transaction $\times$ date fixed effects that soak up potential confounding variation, we interpret this as evidence of a "gender misconduct gap". We conduct two checks below that corroborate the robustness of our design and thus baseline results.

V.2 Falsification Checks

I. Heterogeneity in Transaction Type: As discussed in the Research Design section, our transactional exercises included transactions, OTCs, that are structurally vulnerable to misconduct. One would expect such space of transactions to drive our estimated gender effects since the observed pattern of misconduct holds across gender (see, either Table B.6 or Figure 1). To test for this possibility, we re-estimated Equation (1) by including an interaction with a dummy variable that indicates whether or not the transaction is an OTC. The results are displayed in Table C.1, and provide a strong evidence that the baseline results are driven by over-the-counter transactions. This finding reaffirms our descriptive evidence indicating that misconduct is concentrated in OTC transactions.

II. Re-Assignment of Gender Matches: To what extent are our results on gender misconduct gap driven by the one-time random assignment of customers to vendors? In a follow-up, we re-matched the vendors and customers whereby the initial gender assigned to vendors was reversed. Table C.2 reports the results, and show very similar estimates and inference (with marginal improvements in standard errors, as expected). This suggests that our baseline results on gender misconduct are robust to the gender assignment. In part, this explains why our central results on misconduct gap remain unchanged even after controlling for the gender of customers. The observed patterns of misconduct are unlikely driven by customers' interpersonal variability, which is congruent with the initial evidence of strong covariate balance. Going forward, we include the transaction data from all the audit rounds for more variation.[9]

V.3 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.

I. Evidence of Market Discrimination

Table 3 shows the results from Equation (2). As indicated, $\mathbf{X}_i'$, includes vendors' gender in columns (2)-(3) and (5)-(6). The indicator for Consumer Assignment: Female is large and

[9]In an early draft version of this paper, we used the transaction data from only the first wave, which exludes the data from this follow-up round (wave 2 audit exercise). See Table B.1.

significantly positive across all outcomes and specifications. From the double-post-selection LASSO, we estimate that vendors are about +96% (22 pp) more likely to cheat female customers as compared to similar customers who are males. This corresponds to +89% for the severity outcome, and provides strong evidence that more financial misconduct is committed against female customers once you eliminate endogenous customer-vendor matches.

II. 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 4. Relative to a Male-Male Match, female vendors are more likely (with an estimate of +28 pp) to cheat female customers but +13 pp more likely to cheat similar customers who are males. However, male vendors are 25 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 ($\frac{0.25}{0.23} \times 100 = +108\%$) 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 +15 pp (28 pp-13 pp, respectively) more misconduct of female vendors against female customers. This difference is economically large (about +65%) but not statistically significant at conventional levels (p-value=0.12, significant at the 12% level).

Together, and when combined with the evidence of general discrimination against females (see, Table 3), our results point to misconduct asymmetry: within-gender favoritism for males and within-gender discrimination for females. As we noted earlier, conventional policies aimed at limiting discrimination in organizations and financial markets by increasing the

[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.

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.

III. 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 mobile banking 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]

VI Possible Mechanisms and Discussions

What explains the gender-cheat gap, general discrimination, within-gender discrimination versus within-gender favoritism? To explore this question, we begin with three major competing hypotheses. We then look at unique attributes about female vendors relative to males in this market, and then explore alternative theories that could be at play using several heterogeneity analyses. Lastly, we examine the potential role of differences in perceptions (or beliefs) among the transacting parties (vendors and customers). The results, of which, can help guide policy designs aimed at reducing misconduct and discrimination in markets.

[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 fundamental or official levels.

VI.1 Power: Differences in Women Empowerment

Results on misconduct gap? There are at least two, non-mutually exclusive ways that regional variation in women empowerment can help explain the results on gender misconduct gap. In theory, regions that are characterized by high women empowerment at baseline could create incentives for more gaps. This is plausible because high female empowerment will suggest a lower social cost of misconduct, if caught, for female vendors. Alternatively, regions that are characterized by low women empowerment may create incentives for more gaps. This could occur because female vendors would have to make more profits to be considered good in business or at work, creating incentives for "excessive" profit maximization objectives to become relevant, and thus higher misconduct. Notice that both hypotheses may coexist but the overall outcome depends on the one that is significant.

We draw on data about women empowerment from the most recent Demographic Health Survey (DHS) to explore these two hypotheses. We adapt two common indices of women empowerment (see, eg., 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 experimental 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. Finally, we examine the influence of gender empowerment using a modified version of Equation (1)

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

where Empowered_v is an indicator for localities or vendors in high women empowered districts. Table 5 displays the results, and shows that the gap in misconduct is concentrated in low empowered localities. Thus, we find the most support for low empowerment of females which creates incentives for excessive profit maximization objectives and misconduct of female vendors. This has a simple game-theoretic interpretation: the expected net benefit of cheating is non-nonnegative for low-empowered females to consider misconduct a dominant strategy. Indeed, this is likely to be the case if the females find it either more costly to establish reputational capital or difficult to enter the market.

What of the results on misconduct asymmetry? The gender differences in empowerment also provide a possible explanation. To explore this, we use the following model which interacts the random assignment of customers with female empowerment. 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. First, we estimate a modified version of Equation (2)

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

Table 6 shows the results, and provides evidence that the disproportionate cheat against female customers diminishes under equal or high women empowerment. Second, we estimate the following version of Equation (3)

$$
\begin{aligned}
m_{ivtd} = \ & \gamma_1 \text{Female-Female}_i \times \text{Empowered}_v + \gamma_2 \text{Female-Male}_i \times \text{Empowered}_v \\
& ... + \gamma_3 \text{Male-Female}_i \times \text{Empowered}_v + \beta_1 \text{Female-Female}_i + \beta_2 \text{Female-Male}_i \\
& ... + \beta_3 \text{Male-Female}_i + \mathbf{X}_i' \xi + \eta_v + \mu_{td} + \epsilon_{ivtd}
\end{aligned}
$$

The results are reported in Table 7, and shows an additional strong evidence that the disproportionate cheat over female customers diminishes under equal or high women empowerment. Together, our results indicate that gender differences in empowerment provides a major explanation for the estimated gender gaps and asymmetry in misconduct. In section VI.5, we conduct additional tests that provide micro-evidence re-affirming gender differences in empowerment as a relevant mechanism.

VI.2 Risk Taking: Differences in Risk Attitudes

There is a vast literature showing modest to no gender differences in risk preferences (see, Croson and Gneezy 2009; Charness and Gneezy 2012 for reviews). If female vendors were less risk averse than male vendors in this market, then we might expect the female vendors to take the risk of committing more misconduct relative to the males. We explore potential gender differences in risk attitudes using an investment-based measure of risk aversion (adapted from the design in Gneezy and Potters 1997) which has the appealing property of being easy to understand and thus, suitable for low-income field environments (Charness and Viceisza 2015). The investment-based way of measuring risk aversion is also consistent with our underlying domain of market misconduct whereby the vendors stand the chance of making money (i.e., if they commit misconduct and are not caught) or losing money (i.e., if they are caught and punished e.g., losing their license to operate the M-Money business).

In a follow-up exercise, we re-visited a representative subset of the vendors who were asked to complete an investment task. For this task, each vendor was provisionally endowed with GHS10 and could invest any portion of that amount in a risky asset with a 50% chance of success. The investment paid 2 times the amount invested if successful, but nothing if unsuccessful. The portion not invested is retained by the vendor. We were careful to ensure that the players understood the tasks, chances and payoffs: all of these were verbally explained in detail, and the subjects were asked to repeat that to the experimenter. The experimenter flipped a coin to determine success or failure for these investors (subjects were informed earlier), and then computed their total payoffs which were paid in cash. This investment task provides us with an estimate of risk aversion for each vendor, whereby the higher the investment the less risk averse is the vendor. Results are displayed in Figure C.1, showing a large range of investment choices to the investment task.

We use the two-sample Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1933) to determine if there are any differences in the distribution of risky investments for female versus male vendors. The approximate and exact p-values for the test are 0.986 and 0.962, and thus fail to reject the null hypothesis that the two distributions are equal at the 5% significance level (a simple regression of investments on an indicator for whether the vendor is a female or not provides a positive coefficient and p-value$=0.183$; we do not cluster the standard errors to be able to reject the null more often). Our sample size is small here, and perhaps explains why we find insignificant differences. However, previous research has also found either no significant differences in risk attitudes across gender or (rather) slightly higher risk aversion for females. This suggests that the estimated gender misconduct gap is unlikely driven by differences in risk attitudes.

VI.3 Competition: Female vendors located in less competitive localities – Thus, enjoying monopoly rents?

Competition can put limit on the prices and thus, the extent of misconduct. If female vendors are located in villages that are less competitive compared to male vendors, then it might create room for the female vendors to overcharge or fake the transactional charges. We use market sales information of vendors available from the baseline census to estimate competition. The data on vendor sales is used to construct a Herfindahl-Hirschman index, where a lower (higher) index reflects higher levels of competition (market concentration). Results are displayed in Figure C.2.

The graphical results suggest small gender differences in the levels of competition: larger values (less competition) are more probable for females, which might suggest more misconduct for female vendors relative to males. We formally evaluate if there are any meaningful differences in the distribution of competition, as defined by the Herfindahl-Hirschman index, for female versus male vendors, utilizing the two-sample Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1933) for the two groups. The approximate and exact p-values for the test are 0.504 and 0.504, and thus fail to reject the null hypothesis that the two distributions are equal at the 5% level (a simple regression of Herfindahl-Hirschman index on an indicator for whether the vendor is a female or not provides a p-value$=0.442$; we do not cluster the standard errors to be able to reject the null more often). Similarly, this indicates that the misconduct gap is unlikely driven by gender differences in exposure to competition.

VI.4 Attributes and Heterogeneity

This section looks at several attributes about female vendors relative to male vendors, to explore heterogeneity in gender misconduct gaps and other potential channels that may rationalize the estimated gender differences. First, we estimate the model

$$\text{Vendor: Female}_i = \mathbf{X}_i'\xi + \eta_v + \mu_{td} + \epsilon_{ivtd}$$

where the vector $\mathbf{X}_i'$ houses both demographic and business-wide factors. Results are contained in Table C.3. A few of the attributes are significant across the different specifications: females vendors are characterized by relatively less self-employment, lower experience in doing M-Money business, and slightly larger business sales. Perhaps, because of their slightly lower experience in doing business, female vendors tend to spend long times to complete transactions.

We next examine the influence of these attributes using a modified version of the baseline estimating model (emphasizing attributes that are significant)

$$m_{ivtd} = \text{Vendor: Female}_i \times \mathbf{X}_i'\gamma + \beta\text{Vendor: Female}_i + \mathbf{X}_i'\xi + \eta_v + \mu_{td} + \epsilon_{ivtd}$$

Results are shown in Tables C.4 and C.5. Significant differences in attributes that likely drive the gender misconduct gap are possibly related to *marital status* of vendors (for the demographic attributes) and their *tariff posting* behavior (for the business attributes). We discuss below two theories that relate these two attributes and misconduct: gender differences in *marital status* (i.e., monitoring cost effects) and *tariff posting* (i.e., asymmetric information).

VI.4.1 Asymmetric Information: Differences in Price Transparency

In general, information gaps about the transactional prices between customers and vendors may influence the extent of misconduct. But relative to customers, do female vendors have superior knowledge or any informational advantage about the true transaction charges than their counterpart male vendors? If the latter is true, then one might expect this to rationalize the estimated gender misconduct gap. First, results from the attribute-level analysis show no discernible differences between female and male vendors in terms of posting official tariffs at banking sites. This suggests that gender differences in transparency in prices between customers and vendors should play very little role (see, Table C.4). The evidence is inconsistent with the hypothesis that female vendors may have informational advantage over prices than male vendors, and thus does not provide an explanation for the estimated gender misconduct gap.

Next, we draw on data from the baseline market census to also examine if female vendors have superior knowledge about the true prices compared to male vendors. In a series of tests, vendors were asked to indicate the true charges for two randomly chosen transactions of sizes GHS200 and GHS1200. We were careful to inform the vendors at the beginning that we were not there to perform any actual transactions, but to rather assess their overall knowledge about the market for M-Money. The knowledge tests were also taken towards the end of the surveys. These two features remove any potential incentives for misconduct or overcharge in their answers and provide us an estimate of their knowledge about the true charges.

We find that female vendors know less than the male vendors (but better than all customers): with reference to the "official" charges for the two transactions (GHS200 and GHS1200), female vendors were 59% accurate in their answers, while male vendors were 70% accurate. In a simple regression of an indicator for vendor accuracy on Vendor: Female$_i$, we estimate that female vendors are about 11.4 pp less likely to be correct (p-value <0.01 if we do not cluster the standard errors; p-value =0.114 with a locality-level clustering). This indicates that female vendors do not have superior knowledge about the true prices, perhaps due to their relatively lower experience in doing business (Table C.3). Thus, the gender misconduct gap is unlikely to be explained by such gender differences.

VI.4.2 Differences in Effects of Monitoring

One hypothesis that relates to the gender differences in marital status is that the cost of misconduct may be lower for single vendors since other family dependents are minimally affected – if ever caught, reported to the provider or regulator and punished e.g., losing the license to operate the business of mobile banking. Results in Table C.3 suggest that there are (insignificantly) more single female vendors, which will imply that gender differences in marital status should play very little role. However, if a large (but significant) fraction of female vendors are singles, then one might expect the cost of misconduct to be lower and thus commit more misconduct than the male vendors. The available evidence provides very little support for differences in monitoring costs as a major channel that rationalizes the gender misconduct gap.

A Motivating Theory of Misconduct: Price Transparency versus Monitoring

We formulate a simple model that captures relevant features of price transparency and monitoring effects to illustrate formally how these two could potentially act to affect the incidence of misconduct on the market for M-Money. Details are in Appendix A. In the model, customer i submits a transaction of volume: $t_i \in [0, T]$ to the vendor; the vendor then chooses her strategy for misconduct by trading off the returns from committing a misconduct against the cost of being caught.

The model generates the following two implications that provide additional intuition for why observed differences in price transparency and monitoring effects for females do not reconcile the significant misconduct gap.

Empirical Implication 1. *An increase in monitoring vis-a-vis it's cost on vendor's, if caught, will decrease the incidence of misconduct. Thus, misconduct is likely to be higher for vendors whom the cost of monitoring might be lower.*

Empirical Implication 2. *Increasing price transparency vis-a-vis the fraction of customers informed about the true transaction price decreases the incidence of misconduct. Thus, higher levels of asymmetric information about the true transactional charge between vendor and customers will likely increase misconduct.*

Misconduct is negatively influenced by both the cost imposed on vendors–if reported, monitored and caught of any misconduct, and the transparency in transactional charges between vendors and customers. Our estimated "gender misconduct gap" may reflect both differences in monitoring costs and price transparency. Suppose that the monitoring cost–if caught–is lower for "singles" (unmarried) vendors relative to the married. We estimated the share of single female vendors to be (insignificantly) higher than the share of single male vendors. Thus, the significant gender misconduct gap cannot be rationalized by differences in marital status. Similarly, we present evidence that female vendors are insignificantly more likely to post the official charges at banking sites i.e., no differences in price transparency. The model will suggest that female vendors will commit similar misconduct, which does not rationalize the misconduct gap and asymmetry.

VI.5 Taking Stock of Mechanisms: Discussions and Alternative Explanations

So far, we have explored five competing hypotheses that could rationalize our results on gender differences and asymmetry in misconduct. Our evidence does not support four of these five hypotheses, which include: women being more risk taking, being in more market concentrated villages, having lower market costs of misconduct, and being in less price-transparent villages. However, our evidence provides the most support for differences in social empowerment whereby preexisting low female empowerment incentivizes excessive profit maximization motives for female vendors in the business of mobile banking. As shown in Table C.3, female vendors tend to have slightly larger and significant business sales relative to male vendors, which likely supports the hypothesis of higher profit maximization objectives of female vendors. This creates incentives for more misconduct of the female vendors, which is committed more on female customers than on male customers, who are presumably more empowered than the female vendors. *Mutatis mutandis* male vendors who are also more empowered will cheat the less-empowered female customers.

VI.5.1 Corroborative Evidence on Differences in Empowerment

We have shown that a relevant channel underlying the gender differences in misconduct is low empowerment of female vendors. Since empowerment may come from either wealth (economic status) or leadership status (social standing), the effects would be especially true for very low-income female vendors and localities.

Here we test the potential importance of these dimensions using micro-level data on personal income levels of vendors. In the baseline market census, vendors were asked to indicate their total monthly personal income across five relevant income intervals: less than GHS500, [GHS501-GHS1,000], [GHS1,001-GHS1,500], [GHS1,501-GHS2,000] and above GHS2,000. We convert these increasing income intervals to an ordinal scale of 1 to 5 respectively, and for each locality, we compare the average income score of female vendors to male vendors. If the average score for female vendors is less than the average score for male vendors, we classify the locality as "Less income group". Higher income may strongly correlate with higher female wealth and social status (and thus more women empowerment). Hence, if gender differences in empowerment is indeed a major mechanism for our estimated gender gaps and asymmetry in misconduct, then one might expect the effects to be concentrated in localities that contain female vendors with lower income than their counterpart male vendors (i.e., "Less income group"). We test this by re-estimating Equation (1) separately for markets classified as Less income group versus not.

The results are reported in Table C.6, showing that the gap in misconduct is strongly driven by the markets containing female vendors with relatively lower income. This finding provides micro-evidence that supports the low empowerment of female vendors' channel.

VI.5.2 Evaluation of Alternative Hypotheses

Do differences in empowerment explain all the estimated gaps and asymmetry in misconduct? What of the potential role of gender differences in social cost of misconduct? spatial clustering of female vendors and the implied peer influences? other relevant features of doing business? and differences in perceptions between vendors and customers? In what follows, we discuss the possibilities for alternative channels.

I. Social Cost of Misconduct Lower for Females (If caught)? Our analysis of women empowerment suggests the opposite effect, and thus this cannot significantly explain either the gap or asymmetry in misconduct.

II. Significant Peer Effects along Gender Lines? In general, learning from peers or co-vendors and other influences in a local market could account for the prevalence of financial misconduct (Bursztyn et al. 2014; Dimmock, Gerken and Graham 2018). However, to what extent are these peer effects potentially "gendered" (i.e., peer influences occurring along gender boundaries) and thus influence our results on gender misconduct gap? Data from the local market census allows us to calculate the percentage of female vendors in each locality. Figure C.3 shows the distribution, illustrating a large range of values. We restrict attention to localities with at least 2 vendors (represents 90% of the localities) in calculating this; the overall results are robust to removing this restriction.

We re-estimate Equation (1) by including an interaction for the percent of female vendors in the local market. Results are displayed in Table C.7, and show that all the interaction terms are highly insignificant with occasionally negative signs. Our interpretation is that, peer effects are present, but these effects are not significantly "gendered". Indeed, if peer effects significantly occur along gender lines, then one would possibly expect this to be higher for male vendors (and not females) given that there are more male vendors (60%+ of the market) and relatively more experienced in doing the business of M-Money than females. Finally, notice that these results are neither driven by (i) sample size issues because one can replicate the baseline effects using this sample, nor (ii) model specification(s) because including the direct controls for percent of female vendors in the model does not change the results.

III. Business Dimensions: Female vendors have slightly larger business sizes, and thus might exert some level of non-competitive market power to influence misconduct. However, this effect is unlikely meaningful, as exemplified by the HHI estimates (Figure C.2). Similarly, differences in business experience or formal education do not significantly explain the gap in misconduct. There is no significant difference in formal education between female and male vendors, and the male vendors have slightly higher experience in doing business than female vendors. In addition, the attribute-level analysis shows that differences in either business experience or formal education have limited effects in influencing the estimated misconduct gap.

IV. Misconduct: 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 C.4 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 attribute-level analysis (Table C.3) which shows no significant differences in several relevant vendor dimensions.

V. Differences in Market Beliefs about Misconduct and Gender: Differences in beliefs held among transacting parties may influence the extent and distribution of misconduct. If, for example, customers perceive vendors as fraudulent (which likely decreases vendors' reputation), and there is a negative return to bad reputation, then, in areas where customers believe that vendors commit misconduct, the vendors (i.e., female vendors, who commit misconduct more often) might react by decreasing their level of misconduct. To test this, we re-estimate the misconduct gap interacted with a dummy variable for customers' beliefs about misconduct. In the baseline census, customers were asked if they believe vendors overcharge M-Money and other financial services the vendors offer (about 46.5% of 1,524 subjects indicated `Yes`). We use this to create a simple 0-1 indicator for customers' beliefs.

Results are shown in Table D.1, indicating less misconduct gaps in market areas where customers perceive that vendors commit misconduct (about -21 pp or -GHS0.83 decrease). This provides descriptive evidence that customers' perception about vendor misconduct is important and likely at play.

We look at the role of beliefs further. If, in addition, 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 and other dimensions of beliefs, between April-May 2020, we deployed a phone survey (due to COVID-19 disruptions) of 214 subjects (32 vendors and 182 customers) across 32 localities to gather perceptions of local market participants on various aspects of 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 (i.e., `Agree`/ `Disagree`) and incentivized guess about the percentage of others (all vendors and customers in their locality) that will `Agree` to the statements. Details about the statements are contained in Table D.2 of Appendix C. Survey results are summarized in Figures 2 and D.1.

First, 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.

Second, subjects (both vendors and customers) underperceive the level of misconduct by female vendors relative to male vendors. Just 19% of respondents believe that female vendors are more likely to overcharge customers relative to male vendors but respondents estimate that 40% of others in the local market will agree that female customers are likely to commit more misconduct. In the main experiment, we estimate that female vendors are 10 pp more likely to commit misconduct compared to male vendors. There are no significant gender differences in beliefs for either vendors or customers. Misconduct may be hidden and hard to detect by others e.g., customers. This evidence is consistent with why female vendors act opportunistically and thus commit more misconduct than male vendors.

Third, if we combine the two pieces of evidence, then one would also expect that female vendors commit more misconduct against female customers than male vendors will do against female customers, as found in our experimental trials. But why may male (including female) vendors overcharge male customers? Male customers are generally perceived to be more receptive to transactional overcharge. Respondents estimate that over 55% and 45% of others in the local market will believe that male customers are more receptive to misconduct by female vendors and male vendors, respectively.

Thus, the market's view that male customers are more sophisticated, and customers' under-perception of the misconduct level by female vendors are likely relevant channels that may underly our estimated gender differences in misconduct. Consequently, we hypothesize that differences in beliefs between the transacting parties (vendors and customers) provide an additional rationalization for both our results on gender misconduct gap and asymmetry. Although we find very limited support for other relevant explanations, it is interesting to explore these alternatives and compare them based on gender.

VII 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 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 (+40%) more misconduct relative to their male counterparts. Interestingly, while female vendors discriminate against customers of their gender, male vendors favor customers of their gender. All market vendors, however, cheat female customers (+89% to +96%) more as compared to similar customers who are males.

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), parts of the market that conventional theory e.g., risk preferences, might predict to not hurt is where misconduct actually happens. Second, our results illustrate that gender differences in empowerment is a major explanation for financial misconduct. In addition, we emphasize differences in beliefs about gender and misconduct among transacting parties (vendors and customers) as another parallel explanation. 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 (i.e., preexisting differences in gender empowerment and beliefs about gender) can lead to undesirable market outcomes and may be an important source of financial market frictions. Hence, tackling gender empowerment 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 women empowerment is low, consumer sophistication 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, and has broader implications for the design of innovative financial instruments aimed at influencing financial market development and inclusion in low-income and emerging societies.

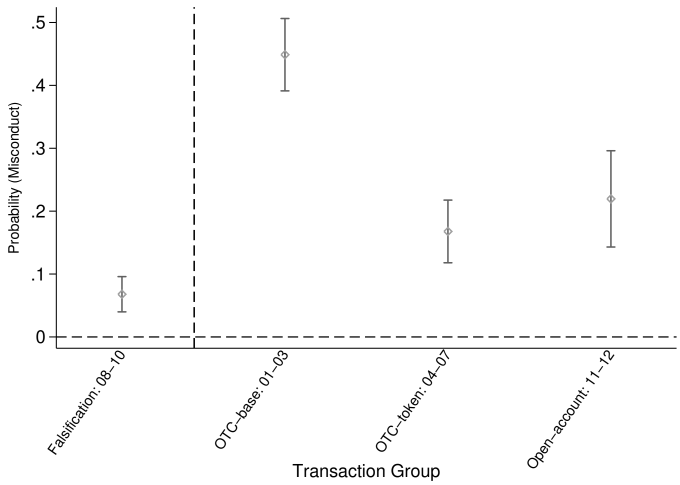
Main Tables and Figures for Text

Table 1: **SUMMARY STATISTICS OF RELEVANT VARIABLES FROM THE MARKET CENSUS**

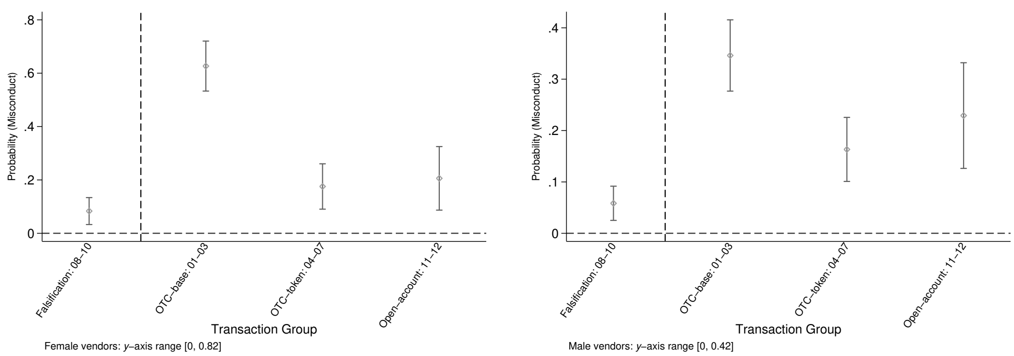
	Vendors		**Customers**	
	Mean	SD	Mean	SD
Female	0.39	0.489	0.62	0.484
Self employment	0.47	0.499	0.68	0.466
Self income intervals [GHS] (monthly)	2.01	1.483	1.37	0.868
Married	0.24	0.432	0.53	0.498
Akan ethnic	0.57	0.494	0.62	0.485
Age (years)	26.2	8.242	39.5	15.02
Education (any)	0.69	0.461	0.89	0.304
M-Money training	0.50	0.500		
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
Household head read English	0.76	0.421	0.60	0.488
Outer wall used cement	0.74	0.433	0.70	0.456
Toilet facility	0.89	0.311	0.84	0.357
Working mobile phone(s)	0.97	0.152	0.97	0.151
Own working bicycle/ motor bicycle/ car	0.28	0.449	0.21	0.410
Market: Features + Transactions + Sales				
Doing business experience (years)	2.05	2.12		
Joint venture: M-Money + other services	0.75	0.431		
M-Money: Total volume [GHS] (daily)	2260	3775		
Non M-Money: Number customers (daily)	32.7	47.06		
Non M-Money: Total volume [GHS] (daily)	155	164.5		
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	

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. The census cover 333 vendors and 1,921 customers or households across a space of 137 villages. The exchange rate during the market census period is US$ 1.0 = GHS 5.12.

Figure 1: **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 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 B.7. 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 2: **GENDER AND MISCONDUCT GAP**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Vendor:	0.091	0.080	0.083	0.308	0.265	0.300
Female (β)	(0.044)	(0.036)	(0.036)	(0.115)	(0.144)	(0.143)
	[0.016, 0.163]	[0.013, 0.133]	[0.022, 0.143]	[0.049, 0.566]	[0.025, 0.504]	[0.058, 0.529]
Fixed Effects	None	District, and Transact×Date		None	District, and Transact×Date	
Controls	None	Vendor	Double-Post LASSO	None	Vendor	Double-Post LASSO
Observations	663	657	657	663	657	657
Mean of Dependent Variable	0.228	0.228	0.228	0.757	0.757	0.757

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, 0-1 indicator for whether auditor is related to the vendor visited or not, and auditor's gender. The double-post LASSO specification in columns (3) and (6) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Table 3: **GENDER AND MISCONDUCT ASYMMETRY – I**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Customer Assignment:	0.223	0.231	0.222	0.668	0.740	0.668
Female (β)	(0.096)	(0.103)	(0.090)	(0.406)	(0.442)	(0.381)
	[0.062, 0.382]	[0.059, 0.403]	[0.073, 0.372]	[-0.006, 1.343]	[0.005, 1.475]	[0.040, 1.296]
Fixed Effects	District, and Transact×Date	District, and Transact×Date		District, and Transact×Date	District, and Transact×Date	
Controls	None	Vendor	Double-Post LASSO	None	Vendor	Double-Post LASSO
Observations	942	936	936	867	861	861
Mean of Dependent Variable	0.228	0.228	0.228	0.757	0.757	0.757

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 specification in columns (3) and (6) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Table 4: **GENDER AND MISCONDUCT ASYMMETRY – II**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Female vendor-	0.279	0.296	0.279	0.842	0.901	0.842
Female customer Match (β_1)	(0.096)	(0.099)	(0.090)	(0.409)	(0.431)	(0.383)
	[0.119, 0.439]	[0.130, 0.461]	[0.129, 0.428]	[0.162, 1.522]	[0.184, 1.616]	[0.212, 1.473]
Female vendor-	0.133	0.143	0.133	0.508	0.496	0.508
Male customer Match (β_2)	(0.049)	(0.051)	(0.046)	(0.174)	(0.182)	(0.163)
	[0.051, 0.215]	[0.057, 0.230]	[0.056, 0.209]	[0.217, 0.798]	[0.193, 0.800]	[0.238, 0.777]
Male vendor-	0.246	0.267	0.246	0.776	0.841	0.776
Female customer Match (β_3)	(0.091)	(0.097)	(0.085)	(0.376)	(0.409)	(0.352)
	[0.094, 0.397]	[0.106, 0.429]	[0.104, 0.387]	[0.151, 1.402]	[0.162, 1.520]	[0.196, 1.357]
Fixed Effects	District, and Transact×Date	District, and Transact×Date		District, and Transact×Date	District, and Transact×Date	
Controls	None	Vendor	Double-Post LASSO	None	Vendor	Double-Post LASSO
Observations	942	936	936	867	861	861
Mean of Dependent Variable	0.228	0.228	0.228	0.757	0.757	0.757
p-value (test: $\beta_1 = \beta_2$)	0.150	0.147	0.124	0.422	0.371	0.389

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 specification in columns (3) and (6) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Table 5: **WOMEN EMPOWERMENT EFFECTS AND MISCONDUCT GAP – I**

	Incidence: **1**(Misconduct=Yes)		Amount-Misconduct, GHS	
	(1)	(2)	(3)	(4)
Vendor:	0.132	0.130	0.423	0.436
Female (β)	(0.041)	(0.037)	(0.148)	(0.134)
	[0.063, 0.201]	[0.068, 0.192]	[0.175, 0.669]	[0.215, 0.656]
x Empowered	-0.152	-0.149	-0.452	-0.454
	(0.072)	(0.065)	(0.213)	(0.196)
	[-0.271, -0.032]	[-0.257, -0.041]	[-0.807, -0.096]	[-0.776, -0.131]
Fixed Effects	District, and Transact×Date		District, and Transact×Date	
Controls	Vendor	Double-Post LASSO	Vendor	Double-Post LASSO
Observations	936	936	861	861
Mean of Dependent Variable	0.228	0.228	0.757	0.757

Note: Table shows the effects of women empowerment on differences in 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 auditor's gender. The double-post LASSO specification in columns (2) and (4) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Table 6: **WOMEN EMPOWERMENT EFFECTS AND MISCONDUCT ASYMMETRY – II**

	Incidence: **1**(Misconduct=Yes)		Amount-Misconduct, GHS	
	(1)	(2)	(3)	(4)
Customer Assignment: Female (β)	0.257	0.222	0.653	0.669
	(0.057)	(0.091)	(0.186)	(0.382)
	[0.162, 0.352]	[0.073, 0.372]	[0.343, 0.962]	[0.040, 1.297]
x Empowered	-0.174	-0.213	-0.627	-0.722
	(0.050)	(0.127)	(0.150)	(0.166)
	[-0.258, -0.091]	[-0.422, -0.004]	[-0.876, -0.377]	[-0.995, -0.449]
Fixed Effects	District, and Transact×Date		District, and Transact×Date	
Controls	Vendor	Double-Post LASSO	Vendor	Double-Post LASSO
Observations	936	936	861	861
Mean of Dependent Variable	0.228	0.228	0.757	0.757

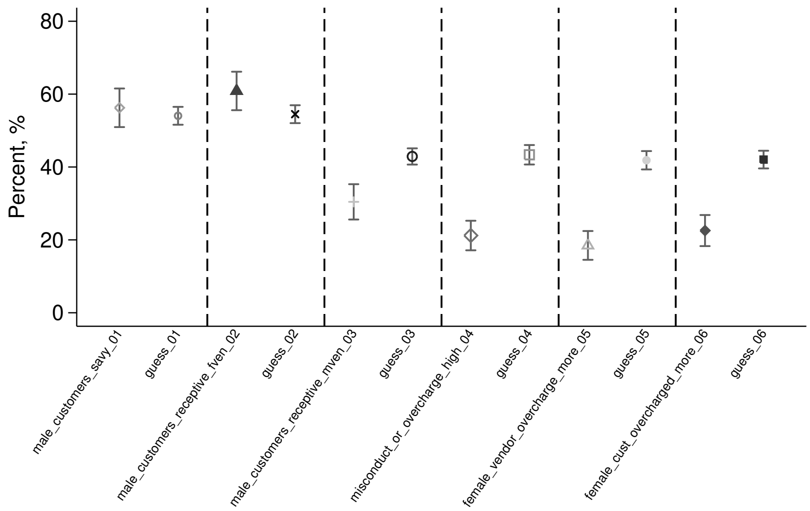
Note: Table shows the effects of women empowerment on the impacts of customers' gender assignment on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a 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 specification in columns (2) and (4) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Table 7: **WOMEN EMPOWERMENT EFFECTS AND MISCONDUCT ASYMMETRY – III**

	Incidence: **1**(Misconduct=Yes)		Amount-Misconduct, GHS	
	(1)	(2)	(3)	(4)
Female vendor-	0.385	0.390	0.978	1.032
Female customer Match (β_1)	(0.081)	(0.069)	(0.240)	(0.194)
	[0.250, 0.520]	[0.275, 0.504]	[0.579, 1.377]	[0.713, 1.351]
x Empowered	-0.262	-0.304	-0.697	-0.771
	(0.080)	(0.073)	(0.242)	(0.214)
	[-0.395, -0.129]	[-0.425, -0.184]	[-1.099, -0.294]	[-1.123, -0.417]
Female vendor-	0.150	0.145	0.574	0.606
Male customer Match (β_2)	(0.053)	(0.047)	(0.178)	(0.166)
	[0.062, 0.238]	[0.066, 0.223]	[0.278, 0.869]	[0.332, 0.879]
x Empowered	NE	NE	NE	NE
Male vendor-	0.281	0.270	0.829	0.830
Female customer Match (β_3)	(0.068)	(0.064)	(0.229)	(0.211)
	[0.167, 0.394]	[0.163, 0.376]	[0.449, 1.210]	[0.482, 1.179]
x Empowered	-0.123	-0.142	-0.544	-0.562
	(0.053)	(0.061)	(0.183)	(0.187)
	[-0,211, -0.034]	[-0.242, -0.041]	[-0.849, -0.239]	[-0.870, -0.254]
Fixed Effects	District, and Transact×Date		District, and Transact×Date	
Controls	Vendor	Double-Post LASSO	Vendor	Double-Post LASSO
Observations	936	936	861	861
Mean of Dependent Variable	0.228	0.228	0.757	0.757

Note: Table shows the effects of women empowerment on 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. 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 specification in columns (2) and (4) 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. NE denotes not estimable, which occurs due to insufficient sample from low women empowerment areas with a male match. Clustered standard errors (at the vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Figure 2: **BELIEFS: ESTIMATE OF MARKET VIEWS AND INCENTIVIZED GUESSES**



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 D.2. 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.

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] Archibong, Belinda, Francis Annan, Anja Benshaul-Tolonen and Oyebola Okunogbe. 2019. "Firm culture: Examining the role of gender and ethnicity in job matching in an online African labor market." Working Paper, Columbia University.

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

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

[8] Banerjee, Abhijit, Marianne Bertrand, Saugato Datta and Sendhil Mullainathan. 2009. "Labor market discrimination in Delhi: Evidence from a field experiment." Journal of Comparative Economics 37 (1): 14-27.

[9] Bannier, Christina E. and Milena Neubert. 2016. "Gender differences in financial risk taking: The role of financial literacy and risk tolerance." Economics Letters, 145 (): 130-135.

[10] 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.

[11] Bernhardt, Arielle, Erica Field, Rohini Pande and Natalia Rigol. 2019. "Household matters: Revisiting the returns to capital among female microentrepreneurs." American Economic Review: Insights, 1 (2): 141-60.

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

[13] 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.

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

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

[16] 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.

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

[18] Charness, Gary and Angelino Viceisza. 2015. "Three risk-elicitation methods in the field: Evidence from rural Senegal." Working Paper, UC Santa Barbara.

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

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

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

[22] Demographic and Health Survey [DHS]. 2014. Technical Report, Ghana.

[23] 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.

[24] Duflo, Esther. 2003. "Grandmothers and granddaughters: Old age pensions and intrahousehold allocation in South Africa." World Bank Economic Review, 17 (1): 1–25.

[25] Egan, Mark, Gregor Matvos and Amit Seru. 2019. "When Harry fired Sally: The double standard in punishing misconduct." Revise and Resubmit, Journal of Political Economy.

[26] Glaeser, Edward L., Bruce Sacerdote and José A. Scheinkman. 1996. "Crime and social interactions." Quarterly Journal of Economics, 111 (2): 507-548.

[27] Glover, Dylan, Amanda Pallais and William Pariente. 2017. "Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores." Quarterly Journal of Economics 132 (3): 1219-60.

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

[29] Goldstein, Itay, Wei Jiang and George Andrew Karolyi. 2019. "To FinTech and Beyond." Review of Financial Studies, 32 (5): 1647–1661.

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

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

[32] Kaufmann D., Kraay A. and Mastruzzi M. 2006. "Measuring corruption: myths and realities." World Bank Policy Research Working Paper.

[33] 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.

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

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

[36] 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.

[37] Sah, Raaj K. 1991. "Social osmosis and patterns of crime." Journal of Political Economy, 99 (6):1272-1295.

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

[39] Sunden, Annika E. and Brian J Surette. 1998. "Gender differences in the allocation of assets in retirement savings plans." American Economic Review, 88 (2): 207–211.

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

Supplementary Appendix

For Online Publication

A A Motivating Theory

We formulate a simple model that captures relevant features of *price transparency* and *monitoring effects* to illustrate how these two candidate explanations could act to affect gender differences and asymmetries in financial misconduct on M-Money. To investigate these effects, one must specify the transactions and charges in the economy.

Setup

We consider a market that is made up of a representative vendor and a set of nearby customers. Consumers are indexed by i , and may submit a transaction of volume: $t_i \in [0, T]$ to the vendor. As we described in the empirical setting, a transaction entails either sending or receiving a cash payment. The vendor will state the charge for this transaction: $c_t = \tilde{c_t} + \varepsilon$, where $\tilde{c_t}$ is the true *per-unit* transaction charge. We assume the following

$$\varepsilon = \begin{cases} 0 & \text{with probability } (1-p) \\ \mu > 0 & \text{with probability } p \end{cases}$$

where ε captures the tendency to overcharge a transaction. This way of defining ε is qualitatively inconsequential to the results even if ε captures genuine mistakes or undercharging. We shall, however, limit our attention to overcharging for the sake of interpreting our empirical results on gender misconduct gaps. The "true" aggregate sales is

$$\int_0^T t_i(1 + \tilde{c_t})di$$

while the "fraudulent" aggregate sales is

$$\int_0^T t_i(1 + \underbrace{\tilde{c_t} + \mu})di$$

We use this framework to first evaluate the potential role of monitoring, and then extend the analysis to allow for the effects of price transparency.

A.1 Effects of Monitoring and Price Transparency

Suppose that some fraction of the *total* transactional volume T customer i submitted to the vendor are likely more monitored. In practice, this may occur if it becomes easy to report any misconduct, or if customers become aware of how to conveniently report vendor misconduct, e.g., decentralizing fraud reporting to the locality level. Denote by $0 \leq s \leq 1$, the fraction of transactions that may be monitored. When monitoring occurs, the transactions are resolved for any glitches and thus $c_t = \tilde{c}_t$. In principle, the customer submits a transaction, and then the vendor decides whether or not to overcharge it. The vendor's problem amounts to choosing p to maximize

$$\max_{0 \leq p \leq 1} \quad T \Pr(c_t > \tilde{c}_t | M = 0) - \underbrace{[T \Pr(c_t > \tilde{c}_t | M = 1)]^2}_{\text{convex costs: } \textit{via} \text{ e.g., reputation loss}}$$

M is an indicator for whether transactions are monitored or not. This shows the tradeoff between the returns from committing a misconduct against the cost of being caught. Solving this problem, we have

$$p^* = \max \left\{ 0, \frac{(1 + \tilde{c}_t + \mu)(1 - s)}{2T(1 + \tilde{c}_t)^2 s^2} \right\}$$

which is decreasing in s.

Proof. See section below.

We now incorporate price transparency. As before, suppose that some fraction of the customers become more financially informed or sophisticated about the "true" transactional charge. We denote this fraction by $0 \leq \theta \leq 1$. Such sophistication may occur if one can reduce the potential information asymmetries about the true transactional price between vendors and customers. Examples will include activities that promote charge transparency: posting of official tariff sheets at banking sites, asking for tariff sheets before carrying out transactions, etc. We allow the extent of mis-charge μ to depend on the fraction of customers who are informed about the true price $\mu(\theta)$, whereby a lower overcharge rate will reflect more charge transparency. Specifically, for $\theta_1 < \theta_2$ this implies $\mu(\theta_1) < \mu(\theta_2)$. We can rewrite the expression for the tendency to mischarge as

$$\varepsilon = \begin{cases} 0 & \text{with probability } (1-p) \\ \mu(\theta) & \text{with probability } p \end{cases}$$

An equivalent solution to the vendor's problem is:

$$p^* = \max\left\{0, \frac{[1+\tilde{c}_t+\mu(\theta)][1-s]}{2T[1+\tilde{c}_t]^2 s^2}\right\}$$

which is also decreasing with θ.

Proof. See section below.

A.2 Results and Implications

We can summarize our analysis using the following proposition, which yields two empirical implications that are congruent with our empirical evidence.

Proposition. *The vendor's strategy for misconduct* (p^*) *is characterized by*

$$p^* = \begin{cases} \max\left\{0, \frac{(1+\tilde{c}_t+\mu)(1-s)}{2T(1+\tilde{c}_t)^2 s^2}\right\} & \text{under monitoring} \\ \max\left\{0, \frac{[1+\tilde{c}_t+\mu(\theta)][1-s]}{2T[1+\tilde{c}_t]^2 s^2}\right\} & \text{under monitoring + price transparency} \end{cases}$$

The probability of overcharging p^* *is increasing in its parameter* μ *and decreasing in the intensity of monitoring* s*. Further,* p^* *is decreasing in the intensity of promoting charge transparency* θ*.*

Proof.

Consider the vendor's objective function

$$\max_{0\leq p\leq 1} \quad T\,\Pr(c_t > \tilde{c}_t|M=0) - \underbrace{[T\,\Pr(c_t > \tilde{c}_t|M=1)]^2}_{\text{convex costs: } via \text{ reputation loss etc...}}$$

The left term captures the benefits of committing a misconduct, which accrues when the transactions are not monitored $M = 0$. $\Pr(c_t > \tilde{c}_t|M = 0)$ is the probability of committing the misconduct that is not detected and returns $(1 + \tilde{c}_t + \mu)$. The right term measures the cost of committing a misconduct that is detected. In this instance, the transactions are monitored $M = 1$, resolved for any mischarges and returns $c_t = \tilde{c}_t$. As is standard in the literature, we allow the cost of a detected misconduct to be a quadratic or convex function of it arguments. The vendor's objective function is now written as

$$\max_{0\leq p\leq 1} \quad T(1+\tilde{c}_t+\mu)(1-s)p - [T(1+\tilde{c}_t)sp]^2.$$

The the first-order-condition is $T(1 + \tilde{c}_t + \mu)(1 - s) = 2[T(1 + \tilde{c}_t)s]^2p^*$. Thus the vendor's strategy for misconduct is chosen to equate the marginal benefits with the marginal costs, respectively. Solving for p^* gives

$$p^* = \max\left\{0, \frac{(1+\tilde{c}+\mu)(1+s)}{2T(1+\tilde{c})^2s^2}\right\}$$

$\blacksquare$

For the case that incorporates price transparency, the proof follows immediately, where

$$p^* = \max\left\{0, \frac{[1+\tilde{c}+\mu(\theta)][1+s]}{2T[1+\tilde{c}]^2s^2}\right\}$$

$\blacksquare$

Empirical Implication 1. *An increase in monitoring vis-a-vis it's cost on vendor's–if caught–will decrease the incidence of financial misconduct. Thus, misconduct is likely to be higher for vendors whom the cost of monitoring might be lower.*

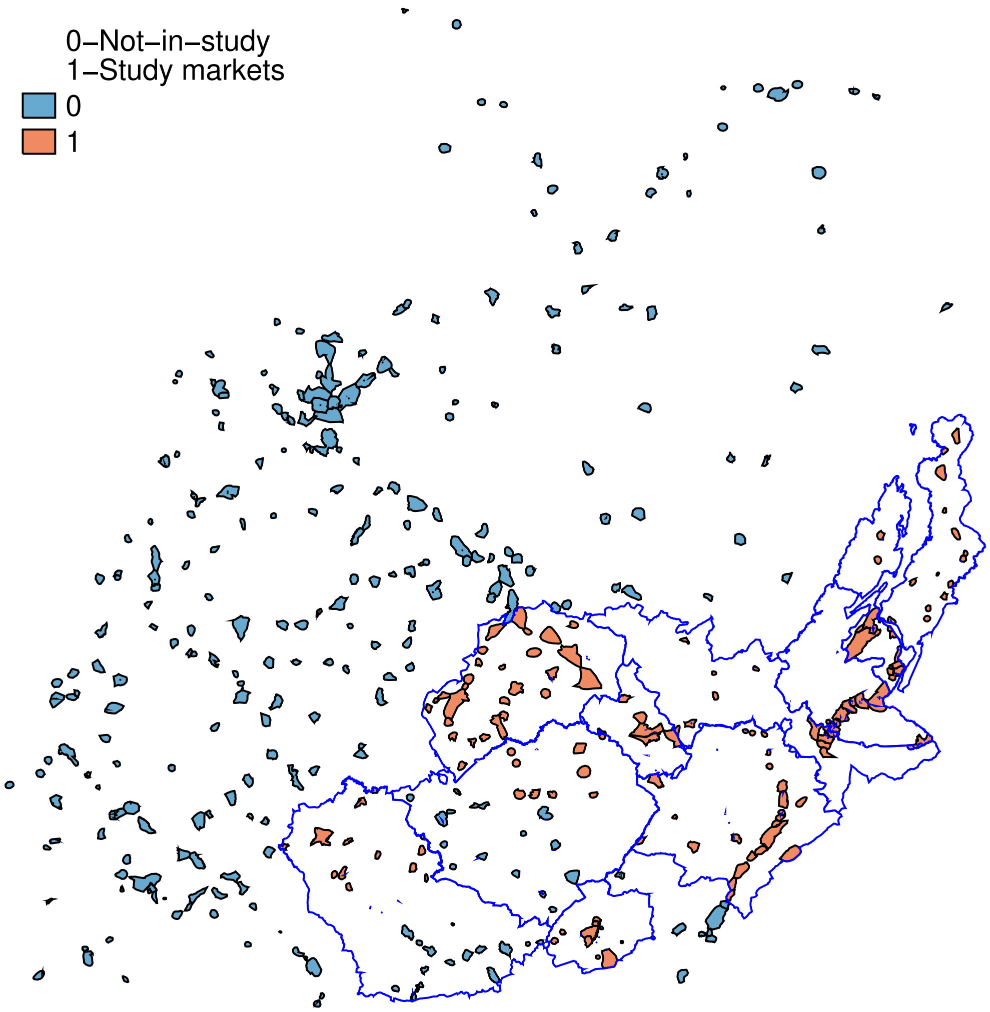
Empirical Implication 2. *Increasing price transparency vis-a-vis the fraction of customers informed about the true transaction price decreases the incidence of financial misconduct. Thus, higher levels of asymmetric information about the true transactional charge between vendor and customers will likely increase misconduct.*

B Further Results: Data and Descriptives

Table B.1: **Study timeline**

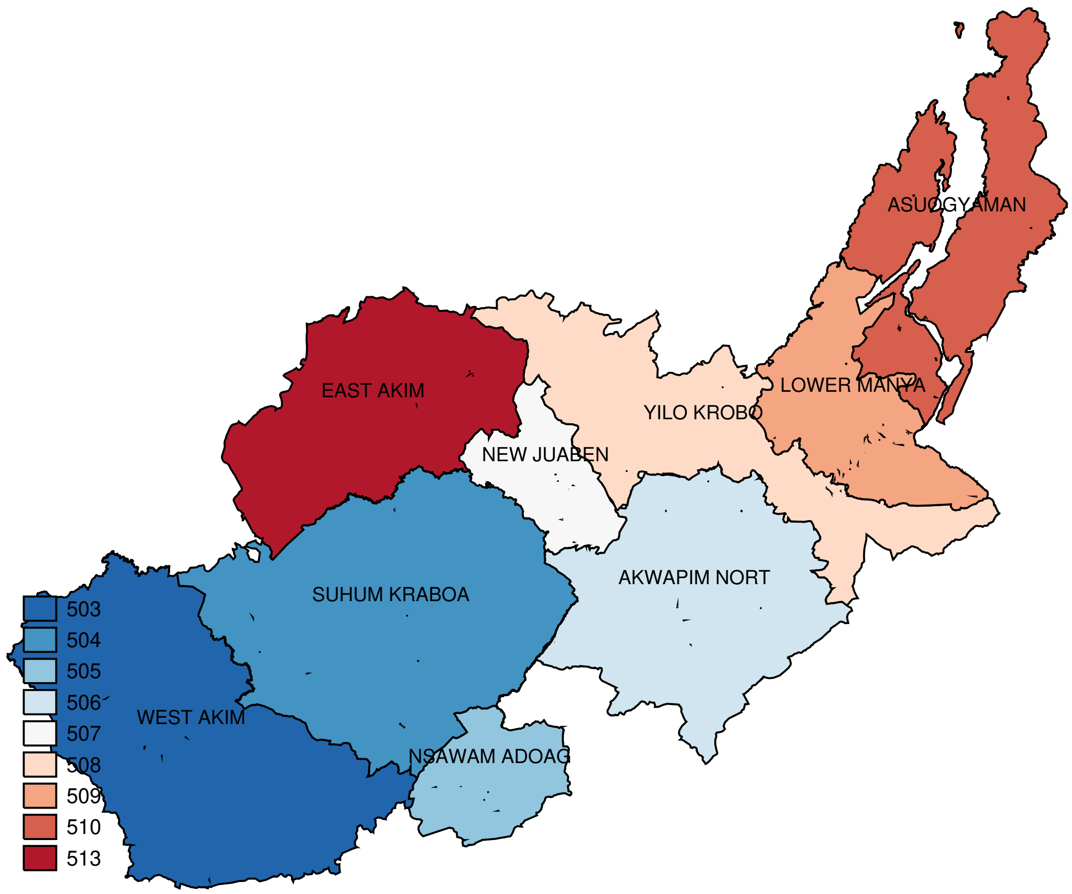
	DATE	ACTIVITY
Part 1	February 2017	**Pilots:** Misconduct – incidence, correlates, design (Annan 2017)
Part 2	February 15-March 20, 2019	**Baseline:** Market census – detail market records
Part 3	August 2019	**Randomization:** Sampling, Auditors recruitment (through field partners, GSS), Experimental customers-vendor assignments + training
Part 4	September 01-October 15, 2019	**Experiment:** Wave 1 audit exercises (main) Wave 2 audit exercises
Part 5	 October 16-October 30, 2019 April 25-May 30, 2020	**Follow up visits – test additional mechanisms:** Risk preferences elicitation, Beliefs about gender and misconduct elicitation

Figure B.1: **Maket 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 market census and subsequent experimental studies. 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 census, 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 B.2: **THE 9 EXPERIMENTAL DISTRICTS**



NOTE: The legend reflects district codes

Table B.2: **SUMMARY STATISTICS OF VENDORS BY GENDER FROM THE MARKET CENSUS**

	Females		**Males**	
	Mean	SD	Mean	SD
Self employment	0.43	0.496	0.50	0.500
Self income intervals [GHS] (monthly)	2.56	1.681	1.69	1.254
Married	0.27	0.445	0.23	0.424
Akan ethnic	0.57	0.495	0.57	0.494
Age (years)	25.7	7.823	26.6	8.493
Education (any)	0.74	0.435	0.65	0.475
M-Money training	0.52	0.499	0.49	0.500
Poverty Indicators				
Household size (above 5)	0.21	0.410	0.23	0.421
Household head read English	0.74	0.434	0.78	0.411
Outer wall used cement	0.74	0.435	0.75	0.432
Toilet facility	0.92	0.268	0.87	0.335
Working mobile phone(s)	1.00	0.000	0.96	0.195
Own working bicycle/ motor bicycle/ car	0.19	0.393	0.34	0.474
Market: Features + Transactions + Sales				
Doing business experience (years)	1.76	1.847	2.24	2.275
Joint venture: M-Money + other services	0.75	0.431	0.75	0.432
M-Money: Total volume [GHS] (daily)	2380	4927	2180	2757
Non M-Money: Number customers (daily)	26.1	26.47	37.2	56.35
Non M-Money: Total volume [GHS] (daily)	136	133.7	167	181.1
Number of observations	140		193	

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

Table B.3: **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 B.4: **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 B.5: **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 B.6: **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 $\times$ 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		

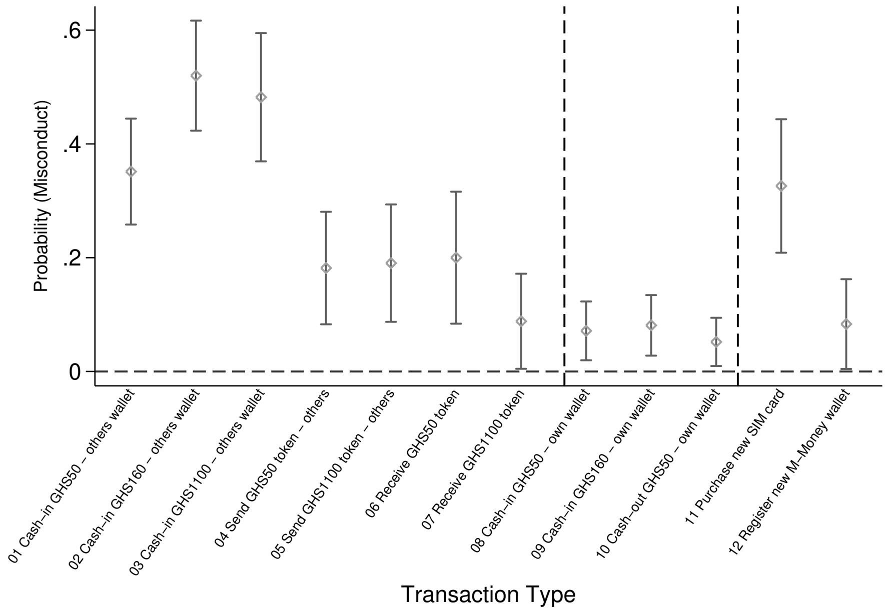
Note: Table shows the descriptive statistics of financial misconduct. These misconduct outcomes are based on the actual transactional exercises. Transactions are categorized into four groups, namely: OTC-base, OTC-token, Falsification, and Open-account. OTC denotes over-the-counter and captures transactions that involve little to no automation or active verification from the side of the customer. The groupings and specific transactions in each transaction group are described in Table B.7. **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 vendor; 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 B.7: **Misconduct: descriptive statistics based on audit transactional exercises – details**

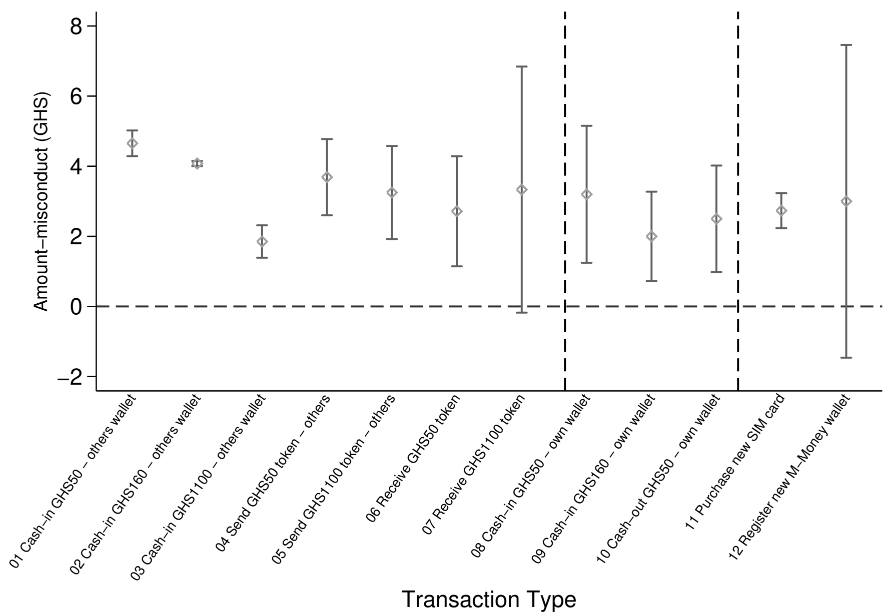
# Transaction type (description)	Outcome variable	Mean	SD	Transaction group	Mean	SD
01 Cash-in GHS50 - to others wallet	**1**[Misconduct=Yes]	0.35	0.480		0.44	0.498
	Overcharged [GHS]	4.65	1.093		3.58	1.498
02 Cash-in GHS160 - to others wallet	**1**[Misconduct=Yes]	0.52	0.502	$\{= \quad OTC-base$		
	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 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].

Figure B.3: **MISCONDUCT: DESCRIPTIVE STATISTICS BASED ON AUDIT TRANSACTIONAL EXERCISES – DETAILS**



(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 transactional exercises. Details of the specific transactions (01-12) are contained in Table B.7. 90% confidence intervals are displayed around the estimates.

C Further Results: Robustness and Mechanisms

Table C.1: **OTC EFFECTS AND MISCONDUCT GAP**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Vendor:	-0.056	-0.009	0.001	-0.215	0.003	0.039
Female (β)	(0.042)	(0.037)	(0.034)	(0.149)	(0.135)	(0.1426)
	[-0.126, 0.013]	[-0.072, 0..052]	[-0.056, 0.057]	[-0.463, 0.031]	[-0.222, 0.229]	[-0.168, 0.247]
x OTC	0.489	0.259	0.254	1.747	0.818	0.786
	(0.068)	(0.095)	(0.087)	(0.295)	(0.384)	(0.350)
	[0.375, 0.603]	[0.101, 0.419]	[0.110, 0.399]	[1.256, 2.238]	[0.180, 1.456]	[0.210, 1.363]
Fixed Effects	None	District, and Transact$\times$Date		None	District, and Transact$\times$Date	
Controls	None	Vendor	Double-Post LASSO	None	Vendor	Double-Post LASSO
Observations	663	657	657	663	657	657
Mean of Dependent Variable	0.228	0.228	0.228	0.757	0.757	0.757

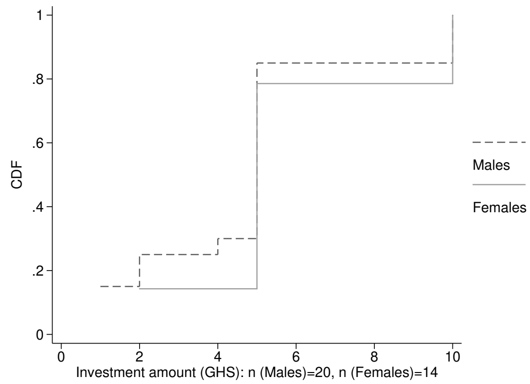
Note: Table shows the effects of vendors' gender on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. OTC is a 0-1 indicator for over-the-counter (i.e., vulnerable) transactions. Direct effect for OTC soaked up in transaction 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 auditor's gender. The double-post LASSO specification in columns (3) and (6) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Table C.2: **RE-ASSIGNMENTS: GENDER AND MISCONDUCT GAP**

	Incidence: **1**(Misconduct=Yes)		Amount-Misconduct, GHS	
	(1)	(2)	(3)	(4)
Vendor:	0.085	0.090	0.278	0.300
Female (β)	(0.033)	(0.031)	(0.118)	(0.112)
	[0.029, 0.141]	[0.034, 0.136]	[0.081, 0.474]	[0.109, 0.480]
Fixed Effects	District, and Transact×Date		District, and Transact×Date	
Controls	Vendor	Double-Post LASSO	Vendor	Double-Post LASSO
Observations	936	936	861	861
Mean of Dependent Variable	0.228	0.228	0.757	0.757

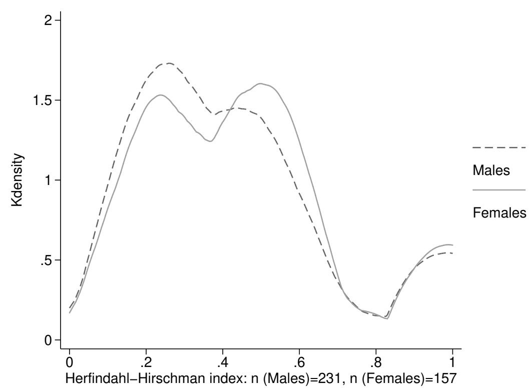
Note: Table shows the effects of vendors' gender on vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. Vendor controls include: 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, 0-1 indicator for whether auditor is related to the vendor visited or not, and auditor's gender. The double-post LASSO specification in columns (2) and (4) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Figure C.1: **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.

Figure C.2: **COMPETITION EXPOSURE BY GENDER**



Notes: Figure shows the distribution of Herfindahl-Hirschman index (HHI) for female and male vendors. A lower (higher) index reflects higher levels of competition (market concentration). The HHI is constructed using market sales information of vendors available from the baseline census, and illustrates the potential gender differences in the levels of competition. There is *limited* graphical evidence that the two distributions are significantly different (consistent with results from a formal two-sample Kolmogorov–Smirnov test of distributional equality). A simple regression of the HHI on an indicator for whether the vendor is a female or not provides a p-value= 0.442, suggesting that the two distributions are not significantly different from each other.

Table C.3: **Differences in vendors' personal and market attributes**

DV: **Vendor: Female**

	(1)	(2)	(3)
Age (years)	0.002	0.002	0.001
	(0.008)	(0.007)	(0.007)
1[Married=Yes]	-0.128	-0.105	-0.125
	(0.143)	(0.134)	(0.127)
1[Akan=Yes]	0.091	-0.005	0.033
	(0.100)	(0.128)	(0.142)
1[Self employed=Yes]	-0.271**	-0.105	-0.088
	(0.110)	(0.101)	(0.105)
Experience in business (years)	-0.048*	-0.027	-0.024
	(0.026)	(0.022)	(0.027)
Business size	0.004	0.001	0.002**
	(0.009)	(0.007)	(0.001)
1[Tariff posted=Yes]	-0.052	0.086	0.063
	(0.084)	(0.078)	(0.085)
1[Joint venture=Yes]	0.016	0.031	0.057
	(0.084)	(0.080)	(0.096)
Transaction duration (minutes)	0.011	0.016***	0.020***
	(0.007)	(0.006)	(0.007)
Fixed Effects	None	District	District, and Transact×Date
Observations	936	936	936
Mean of Dependent Variable	0.363	0.363	0.363

Note: Table shows the gender differences in personal and market attributes of vendors who participated in the audit exercises. Vendor: Female is a dummy variable indicating whether or not the vendor is a female at date t. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Observations are at the vendor × date level (and × transaction level for the transaction duration variable) over the period 9/2019-10/2019. Clustered standard errors (at the vendor level) are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table C.4: **HETEROGENEITY: GENDER AND MISCONDUCT GAP BY VENDORS' DEMOGRAPHIC ATTRIBUTES**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Vendor:	0.143	0.013	-0.021	-0.183	-0.422	-0.215
Female (β)	(0.171)	(0.174)	(0.166)	(0.790)	(0.730)	(0.642)
x Age	0.003	0.006	0.007	0.026	0.030	0.027
	(0.004)	(0.004)	(0.004)	(0.021)	(0.020)	(0.019)
x 1[Married=Yes]	-0.190*	-0.107	-0.071	-0.251	-0.108	-0.222
	(0.108)	(0.105)	(0.102)	(0.420)	(0.365)	(0.332)
x 1[Self employed=Yes]	-0.107	-0.123	-0.112	-0.110	-0.121	-0.149
	(0.086)	(0.082)	(0.085)	(0.275)	(0.245)	(0.257)
Age	0.003	-.0005	-0.002	-0.000	-0.010	-0.013
	(0.003)	(0.002)	(0.003)	(0.014)	(0.013)	(0.014)
1[Married=Yes]	0.005	-0.007	-0.016	-0.158	-0.125	-0.051
	(0.069)	(0.070)	(0.069)	(0.315)	(0.243)	(0.216)
1[Self employed=Yes]	-0.004	0.022	0.042	-0.210	-0.041	0.062
	(0.043)	(0.045)	(0.049)	(0.141)	(0.136)	(0.139)
Fixed Effects	None	District	District, and Transact×Date	None	District	District, and Transact×Date
Observations	936	936	936	861	861	861
Mean of Dep. Variable	0.228	0.228	0.228	0.757	0.757	0.757

Note: Table shows the effects of vendors' demographic attributes on differences in 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. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Observations are at the vendor × transaction × date level over the period 9/2019-10/2019. Clustered standard errors (at the vendor level) are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table C.5: **HETEROGENEITY: GENDER AND MISCONDUCT GAP BY VENDORS' BUSINESS ATTRIBUTES**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Vendor:	0.010	-0.015	0.068	0.330	0.307	0.249
Female (β)	(0.082)	(0.073)	(0.078)	(0.270)	(0.224)	(0.279)
x Experience in business (years)	0.029*	0.021	0.020	0.059	0.028	0.046
	(0.015)	(0.013)	(0.015)	(0.052)	(0.049)	(0.052)
x Business size	0.001	0.001	0.002	0.003	0.002	0.001
	(0.001)	(0.009)	(0.008)	(0.003)	(0.003)	(0.003)
x 1[Tariff posted=Yes]	-0.121	-0.133*	-0.130	-0.448*	-0.572**	-0.465
	(0.078)	(0.072)	(0.098)	(0.247)	(0.231)	(0.304)
x 1[Joint venture=Yes]	0.028	0.058	0.037	0.299	0.256	0.111
	(0.078)	(0.073)	(0.094)	(0.292)	(0.262)	(0.284)
x Transaction duration (minutes)	-0.007	0.008	-0.005	-0.121**	-0.065	-0.011
	(0.021)	(0.020)	(0.023)	(0.053)	(0.051)	(0.071)
Experience in business (years)	0.006	0.002	0.002	0.024	0.015	-0.005
	(0.008)	(0.008)	(0.009)	(0.030)	(0.023)	(0.029)
Business size	-0.007	-0.006	0.002	-0.001	-0.001	0.009
	(0.010)	(0.009)	(0.008)	(0.002)	(0.003)	(0.030)
1[Tariff posted=Yes]	-0.068	-0.041	-0.059	-0.050	0.138	-0.011
	(0.049)	(0.049)	(0.054)	(0.129)	(0.119)	(0.156)
1[Joint venture=Yes]	-0.019	-0.024	0.019	-0.240	-0.132	0.116
	(0.044)	(0.048)	(0.058)	(0.145)	(0.131)	(0.160)
Transaction duration (minutes)	0.024	0.006	0.010	0.183***	0.113***	0.003
	(0.016)	(0.017)	(0.018)	(0.037)	(0.041)	(0.063)
Fixed Effects	None	District	District, and Transact×Date	None	District	District, and Transact×Date
Observations	936	936	936	861	861	861
Mean of Dependent Variable	0.228	0.228	0.228	0.757	0.757	0.757

Note: Table shows the effects of vendors' demographic attributes on differences in 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. **1**[.] is a logical indicator that takes the value 1 whenever the argument in the bracket is true, and zero otherwise. Observations are at the vendor × transaction × date level over the period 9/2019-10/2019. Clustered standard errors (at the vendor level) are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table C.6: **INCOME EFFECTS AND MISCONDUCT GAP**

	Incidence: **1**(Misconduct=Yes)		Amount-Misconduct, GHS	
	(1)	(2)	(3)	(4)
	Less income group =No	Less income group = Yes	Less income group = No	Less income group = Yes
Vendor:	-0.045	0.117	0.120	0.294
Female (β)	(0.064)	(0.052)	(0.237)	(0.177)
	[-0.153, 0.062]	[0.027, 0.205]	[-0.276, 0.517]	[0.002, 0.656]
Observations	530	412	484	383
Mean of Dependent Variable	0.228	0.228	0.757	0.757

Note: Table shows the effects of women income status on differences in 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. Less income group=Yes if the average income of female vendors is less than the average income of male vendors in a locality. Observations are at the vendor $\times$ transaction $\times$ date level over the period 9/2019-10/2019. Clustered standard errors (at the vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets. Correlation between Less income group and Empowered is -0.205 (p-value=0.000), suggesting higher female vendors income positively correlate higher women empowerment, as expected.

Table C.7: **GENDERED PEER EFFECTS AND MISCONDUCT GAP**

Incidence: **1**(Misconduct=Yes)

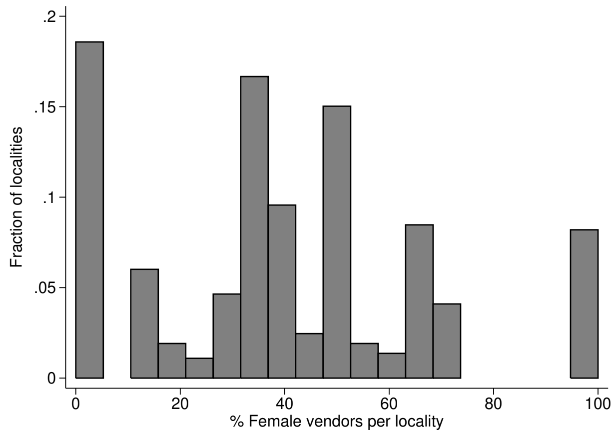
	(1)	(2)	(3)	(4)	(5)
Vendor:	0.019	0.052	0.085	0.127	0.085
Female (β)	(0.077)	(0.064)	(0.057)	(0.068)	(0.054)
	[-0.109, 0.147]	[-0.557, 0.160]	[-0.010, 0.181]	[-0.013, 0.241]	[-0.003, 0.175]
x % Females in local market	0.0008	0.0005	0.0001	-0.0005	0.0001
	(0.001)	(0.104)	(0.0009)	(0.001)	(0.0009)
	[-0.001, 0.002]	[-0.001, 0.002]	[-0.001, 0.001]	[-0.002, 0.001]	[-0.001, 0.002]
Fixed Effects	None	Transact×Date	District, and Transact×Date	District, and Transact×Date	
Controls	None	None	None	Vendor	Double-Post LASSO
Observations	697	697	697	697	697
Mean of Dependent Variable	0.228	0.228	0.228	0.228	0.228

Amount-Misconduct, GHS

	(1)	(2)	(3)	(4)	(5)
Vendor:	0.164	0.129	0.168	0.159	0.168
Female (β)	(0.266)	(0.219)	(0.211)	(0.233)	(0.198)
	[-0.279, 0.607]	[-0.236, 0.495]	[-0.182, 0.520]	[-0.225, 0.543]	[-0.156, 0.494]
x % Females in local market	0.002	0.004	0.002	0.002	0.002
	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
	[-0.004, 0.009]	[-0.001, 0.011]	[-0.004, 0.009]	[-0.004, 0.009]	[-0.003, 0.008]
Fixed Effects	None	Transact×Date	District, and Transact×Date	District, and Transact×Date	
Controls	None	None	None	Vendor	Double-Post LASSO
Observations	641	641	641	641	641
Mean of Dependent Variable	0.228	0.228	0.757	0.757	0.757

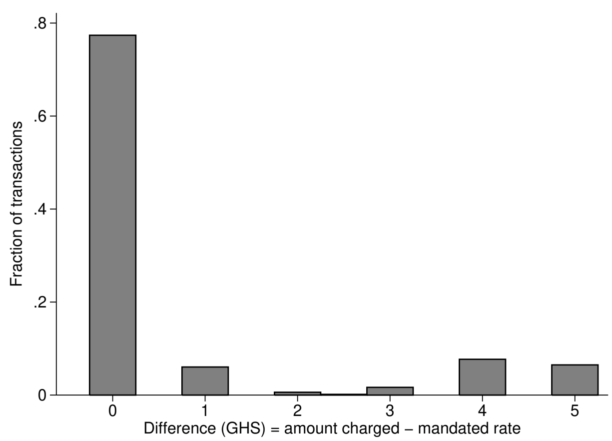
Note: Table shows the effects of gender-based peer effects on differences in vendor misconduct. Incidence is a dummy variable indicating whether or not the vendor committed a misconduct at date t. Amount-Misconduct is the amount (in GHS) overcharged and paid to the vendor as a result of a misconduct. 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 auditor's gender. The double-post LASSO specification in column (5) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Figure C.3: **DISTRIBUTION OF PERCENT FEMALE VENDORS PER LOCALITY**



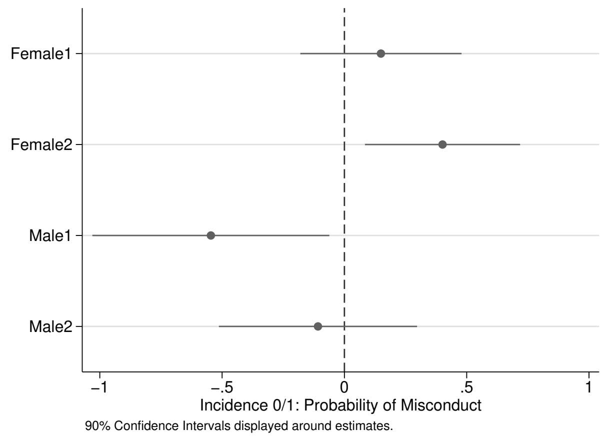
Note: Figure shows the distribution of the % (percentage) of female vendors in each locality. This is calculated using data from the baseline local market census, and illustrates a large range of values: median=38%, mean=39% and standard deviation=28%.

Figure C.4: **DISTRIBUTION OF DIFFERENCE BETWEEN OBSERVED CHARGES AND MANDATED RATES**



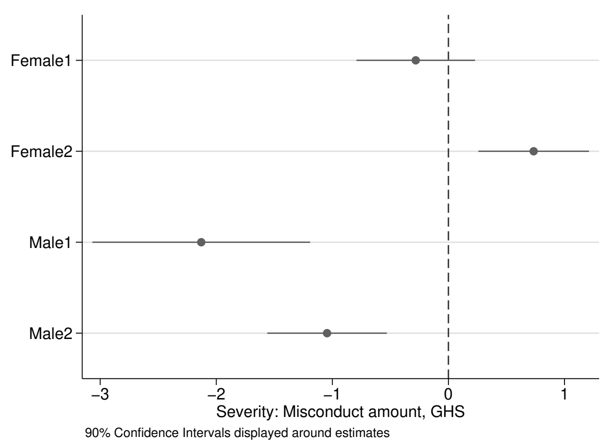
Note: Figure shows the distribution of actual transactional charges relative to the mandated rates. This measures the likelihood of under-charging (if the difference is negative), correct-charging (if the difference equal to 0), and over-charging (if the difference is positive). Calculations are based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

Figure C.5: **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 C.6: **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).

D Elicitation: Beliefs about Gender

Table D.1: **BELIEFS ABOUT MISCONDUCT EFFECTS AND MISCONDUCT GAP**

	Incidence: **1**(Misconduct=Yes)			Amount-Misconduct, GHS		
	(1)	(2)	(3)	(4)	(5)	(6)
Vendor:	0.090	0.115	0.110	0.277	0.419	0.402
Female (β)	(0.033)	(0.041)	(0.039)	(0.118)	(0.166)	(0.156)
	[0.029, 0.141]	[0.046, 0.183]	[0.044, 0.175]	[0.081, 0.474]	[0.44, 0.695]	[0.145, 0.659]
x Believe Misconduct		-0.225	-0.213		-0.881	-0.828
		(0.112)	(0.108)		(0.360)	(0.371)
		[-0.412, -0.037]	[-0.390, -0.034]		[-1.479 -0.283]	[-1.438, -0.217]
Believe Misconduct		0.152	0.141		0.513	0.490
		(0.087)	(0.074)		(0.236)	(0.209)
		[0.007, 0.296]	[0.019, 0.263]		[0.121, 0.9106]	[0.146, 0.834]
Fixed Effects	District, and Transact×Date	District, and Transact×Date		District, and Transact×Date	District, and Transact×Date	
Controls	Vendor	Vendor	Double-Post LASSO	Vendor	Vendor	Double-Post LASSO
Observations	936	647	647	861	647	647
Mean of Dependent Variable	0.228	0.228	0.228	0.757	0.757	0.757

Note: Table shows the effects of customers' beliefs about misconduct on differences in 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. Believe Misconduct is a 0-1 indicator for whether a customer believe at baseline that vendors overcharge financial transactions and has previous experience of 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 auditor's gender. The double-post LASSO specification in columns (3) and (6) 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 vendor level) are reported in parentheses. 90% confidence intervals are reported in brackets.

Market Beliefs about Misconduct and Gender

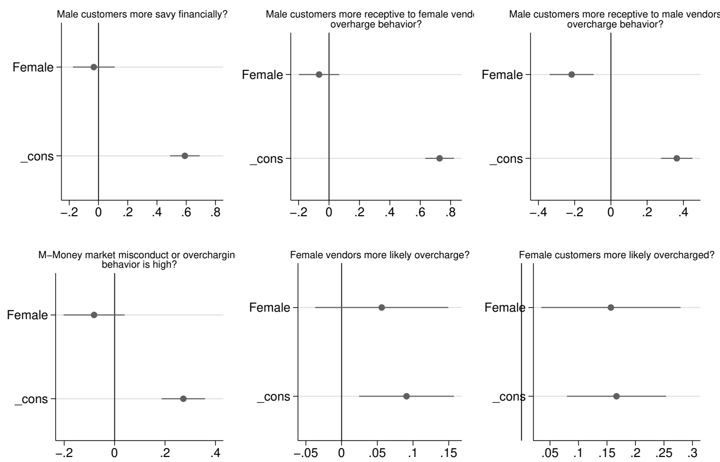
Between April-May 2020, we conducted a wave of phone survey (due to COVID-19 disruptions) to elicit market beliefs, capturing perceptions about various aspects of misconduct on M-Money. For each of the 6 statements below, market participants were asked to indicate their belief (i.e., `Agree/ Disagree`). The respondents consist of a representative sample of 32 local markets: 32 vendors and 182 nearby customers (drawn from our baseline market census). The statements were designed to reflect the gender-differentiated market facts obtained from the main 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 D.2 outlines the specific statements.

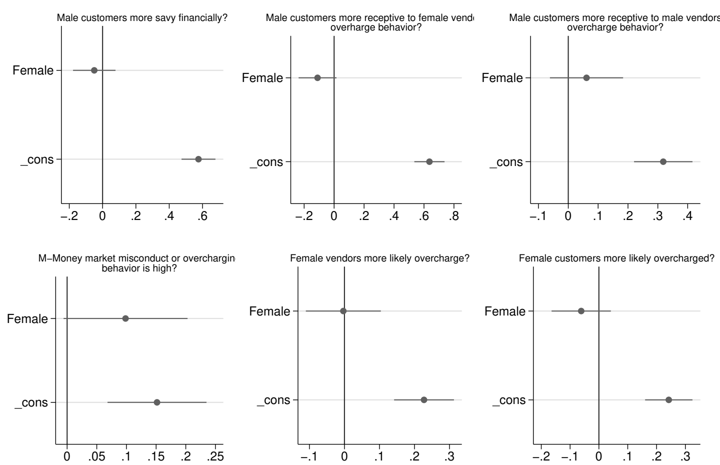
Table D.2: **Belief statements about market misconduct**

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 D.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 D.2. 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.

E 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 on 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 recipients details`
- `Thank you for your service`

**AFTER VISIT: Immediately complete the questionnaire (see, Table E.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 E.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														
.														
.														
.														
.														
.														
.														

. . . . . .