Misconduct and Reputation under Imperfect Information

Francis Annan♣
Georgia State University

v: NOVEMBER 2020

Abstract

Misconduct – market actions that are unethical and indicative of fraud or wrongdoing – is a significant yet poorly understood issue that underlies many economic and financial transactions. Does misconduct in markets matter? When and how does reputation acts as a discipline against market misconduct? We design a field experiment to study the impact of two-sided anti-misconduct information programs on markets, which we deploy on the local markets for mobile money (Human ATMs) in Ghana. We show that, at baseline, these markets are characterized by substantial imperfect information, consumer mistrust, and vendor misconduct. The information programs led to a large reduction in misconduct (-21 pp = -72%) and as a result, an increase in overall market activity, firm sales and consumer welfare. We develop a signaling framework between vendors and consumers that shows the treatment effect is due to a combination of more accurate consumers' beliefs about misconduct and increased reputation concerns. Together, our results indicate a potentially significant source of local financial market frictions, where market activities are underprovided due to misconduct and difficulty in building reputation. Social sanctions through reputational impacts can promote formal local markets when formal sanctions are weak.

Keywords: *forensics and information* (D18, **D83**), *reputation* (**L14**, Z13), *household finance* (D14, **O12**)

♣ Georgia State University, J. Mack Robinson College of Business, 35 Broad St NW, Atlanta, GA 30303. e-mail: `fannan@gsu.edu`. For helpful discussions and suggestions, we thank Arun Chandrasekhar, Xavier Gine, Paul Goldsmith-Pinkham, Bikramaditya Datta, Glenn Harrison, Sean Higgins, Erzo F.P. Luttmer, and seminar participants. We thank Junior Household Finance Seminar, and Virtual CEAR/RMI conference participants for comments on a draft of this paper. Field support from Samuel Adotevi and Kwamena Arkafra (of Ghana Statistical Service) is acknowledged. We are 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. Recent studies have begun to illuminate the nature and potential welfare consequences of misconduct in transactional markets (see, Egan, Matvos and Seru 2019a, Annan 2020; Egan, Matvos and Seru 2019b, respectively). In theory, concerns for reputation (by profit maximizing firms or businesses) deter misconduct and encourage quality provision in markets (Karpoff 2012; Shapiro [1982, 1983]). In practice, however, it might be difficult to establish reputational capital in a market environment with much imperfect information (see, Bai 2019), as it is difficult to demonstrate the quality of services or transactions between transacting parties.

Reputation itself becomes effective and disciplinary *if* there is a high probability of detecting misconduct (Burkhardt 2018) and thus, the presence of imperfect information might exacerbate misconduct, with potential impacts on market efficiency. For example, misconduct can raise the marginal cost of transactions and increase uncertainty in prices (Olken and Pande 2012), which may discourage market demand (Shapiro 1983; Coppejans et al. 2007) and overall firm growth (Jensen and Miller 2018). In emerging and developing financial market contexts, misconduct is likely to be particularly significant as "some" transacting parties are likely uninformed and institutions are weak.

This paper studies such reputation-driven misconduct, addressing three major questions: (i) does financial misconduct impair market efficiency? (ii) to what extent do market parties incorporate symmetric market information to reduce misconduct as a result of reputational concerns? (iii) does such information deepens the market and improves consumer welfare? We report the first study that examines the impact of providing two-sided (market-level) information about market transactions on misconduct, market activity and consumer welfare in Ghana. We conduct a field experiment, drawing on the market for mobile money (aka mobile banking) – a recent financial innovation and well celebrated example of Fintech-based market in developing countries. We construct a unique census of local markets for mobile banking and then perform our experiment by randomly assigning these markets to two candidate anti-misconduct information programs about either price transparency "what to ask while at banking points", monitor and report "how to report transactional glitches or misconduct", or both (their interaction) and then measuring how these affect misconduct, consumers usage and savings on mobile banking with impacts on consumer welfare (risk mitigation and poverty), and vendors sales volume. The different information programs about misconduct provides a means of differentiating between what information is necessary and what is sufficient.

Mobile Money (M-Money) provides financial services and transactions which are delivered on digital mobile networks to potential consumers, 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. Transactional charges and practices are "officially" 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 (as in Egan, Matvos and Seru 2019b; Annan 2020). Here, this captures over-charging and faking transactions with reference to the regulator and provider-approved charges and practices.

The market for mobile banking provides a particularly unique and interesting setting to study misconduct in markets and its impact on households and market activity. First, it is a relatively emerging and growing market with arguably more vulnerable parties and the vast majority of communities have convenient access as compared to, e.g., access to formal banks. The average distance to the closest mobile banking site is 61 meters compared to 339 meters for formal banking versus 383 meters for post office. This alleviates access constraints and allows us to "usefully" examine market behavior and outcomes. Second, our baseline trials indicate (i) significant incidence of misconduct on M-Money (22% of transactions via an administrative audit exercise versus 19-59% via a survey elicitation), (ii) much consumer distrust in transacting at vendor points (62% of customers), and (iii) much imperfect information: random knowledge tests indicate that consumers are less informed about the transactional process and tariffs. Referencing the official charges for two common transactions, we find that customers are 1.5× more likely to be incorrect about the true tariffs relative to vendors who are usually trained prior to the establishment of their businesses – yet have imperfect knowledge about the true tariffs. Relatively higher incidence of misconduct, customer distrust, and poor consumer knowledge imply that the provision of relevant transactional information on misconduct is likely to raise concerns for reputation and be an impactful treatment for the local markets. Together, therefore, our results will likely be relevant for other settings where consumer sophistication is low and financial technology is growing; for example, other sub-Saharan African countries and the Global South.

The experiment involves 130 independent local markets in 130 different localities in eastern Ghana. The large number of markets allows for randomization at the market-level. Markets designate reconstructed pairs of representative vendors and their nearby customers, randomized into the 2×2 information design. The intervention randomly provided the various information sets to both sides of the market (vendors and consumers) and lasted over twenty-two weeks. We tracked several outcomes (at endline): customers or households transaction records, shocks exposure and mitigation, poverty, and collected vendors sales records to examine the supply side effects and to directly validate the household transaction data. We propose a nonstandard audit study to measure vendors' misconduct: trained auditors visited vendor points to make actual transactions, whose charges are benchmarked with the official tariffs to infer misconduct. 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, Ortmann and Hertwig 2002; Kessler, Low and Sullivan 2019). Our main dataset is unique due to its size (i.e., 130 vendors and 990 customers), the expansive set of outcomes from "both" sides of the market, the administrative audit measures of misconduct, and the 2×2 random variation at market-level.

We find four set of results:

- (i) Market misconduct reduced dramatically: The incidence of vendor misconduct decreased as a result of the information programs (-21 pp = -72%+). The joint and monitor and report-only information interventions show economically larger reduction in market vendors' misconduct, however the price transparency-only program also had a negative impact on misconduct.

- (ii) Consumer or household outcomes improved (except for overall poverty): Customers meaningfully increase their uptake of transactional services (0.402 semi-elasticity =+GHS95 per week =+11.2%) and savings behavior (7.6 pp =+12.6%) at vendor points to levels that enable them to better mitigate unexpected shocks (-6.8 pp =-7.6%). The effects on risk mitigation demonstrate a large and objective proxy for resilience and insurance value (Dupas and Robinson [2013]; Breza and Chandrasekhar [2019]) of reducing market misconduct to consumers. We do not find evidence for an impact on overall poverty levels–an objective measure of consumer welfare, which we hypothesize (and verify) that may be distributional (e.g., effects may concentrate on female customers). The joint information program shows larger impacts across the various outcomes, compared to the alternative individual information, suggesting that the two individual information sets are (informationally) complements. Moreover, price-transparency alone (a popular consumer protection instrument) may not be sufficient except when combined with random information assignment about monitoring and reporting.

- (iii) Vendors transactional sales volume increased: Overall, the information programs significantly increased vendors sales volume (+GHS557 per day =+46.5%). This reaffirms the estimated impacts on consumers, and supports the hypothesis that reducing market misconduct enhances the efficiency of local financial markets by increasing the provision of market activity.

- (iv) Closing pre-existing gender gaps: While overall poverty levels did not decrease, it significantly decreased for female customers, particularly for the monitor and report information set (-7.0%). This effect is not significantly different from the joint program's effect. This is interesting because our intervention did not target female consumers, and thus perhaps targeting female customers will likely yield larger impacts. We show that this gender-differentiated impact is explained by the closure of *pre*-intervention gender gaps in usage and savings on M-Money as a result of the anti-misconduct intervention, and argue such elimination of existing gender gaps in uptake and savings behavior improves (gender) equity.

We show robustness of the various findings to several inference procedures, post-double-selection LASSO estimation procedure (Belloni et al. [2014]), including adjustments for multiple testing (List, Shaikh and Xu [2019]) and attrition (Lee [2009], Behaghel et al. [2015]).

What explains the estimated anti-misconduct information impacts? Our underlying hypothesis is that of "reputational concerns", which we also provide evidence for. We set up a simple signaling and reputation framework which guides our information programs and illustrate a reputation interpretation of the results. By providing symmetric market information about vendor misconduct to (potentially uninformed) consumers, it raises vendor concerns for reputation as customers are likely able to infer irresponsible vendors and then assign reputational payoffs to the vendors. *If* vendors care about such (negative or positive) perceptions, then misconduct will fall, which has market-wide implications for the outcomes of our study. The model generates testable implications and allows us to make progress towards the measurement of reputational concerns, which we find evidence for. Our model is an instance of standard microeconomic analysis as applied to misconduct and market behavior, yet our empirical work is innovative: (i) reducing vendor misconduct using a two-sided symmetric market information program, and (ii) measuring reputational concerns based on how customers are able to infer vendors misconduct, and how vendors are able to infer informed or sophisticated customers.

We make three main contributions to the literature. First, is the literature on forensic economics (see e.g., Olken and Pande 2012; Zitzewitz 2012 detail reviews). Misconduct underlies many economic and financial transactions (Egan, Matvos and Seru 2019a, Annan 2020; Egan, Matvos and Seru 2019b), yet the sources of such concealed behavior remain less understood. We emphasize how the presence of imperfect information might exacerbate misconduct in markets, showing in an experiment that providing symmetric information to transacting parties raises concerns for reputation. Very little is known about how reputational looses act as a discipline against business misconduct (Karpoff 2012 provides a review indicating ambiguous effects). In addition, this result speak to the broader notion on the use of (local) sanctions via reputation-building to promote rural financial institutions and development in low-income settings (see, Munshi 2014 for a review).

Second, is the literature on household finance and Fintech adoption (see e.g., Higgins 2020 and references therein). Much research exist on the consumer effects of Fintech, but very little is known about supply side behavior. Here, we emphasize misconduct as key barrier to both sides of the market and that reducing it has meaningful impacts on consumers and vendors. Third, is the literature on information, firm behavior and growth in developing countries (see e.g., Jensen and Miller 2018; Bai 2019). We show that promoting market transparency and monitoring induce firms to seek desirable choices "or quality" with impacts on both efficiency and equity in transactional markets. Firms actions are rewarded via increased consumer demand *if* customers are informed. This increases the size or sales of M-Money firms (that are small to medium sized). Overall, our findings shed lights on why small to medium firms may not grow because they fail to provide quality (by engaging in market misconduct) as quality provision is under-rewarded due to imperfect information.

From a policy perspective, our results highlight how the simple provision of low cost two-sided information might influence misconduct and market beliefs, and how this might eventually facilitate efficient market behavior, particularly in vulnerable market environments. This is important for setting relevant consumer protection policies. Evaluating how uninformed local market subjects are and providing information about price transparency and monitoring (to both sides of the market) could potentially be used to build trust and maximize the benefits of emerging financial markets, particularly Fintech.

II Experiment: Design

II.1 Background: Mobile Money, Market Census, and Market Facts

II.1.1 Mobile Money

The market for M-Money comprises (i) vendors, (ii) customers, and (iii) service providers. Market 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. Particularly, vendors register accounts for customers and act as cash-in and cash-out transaction or banking points for customers (i.e., Human ATMs). 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 has 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 the policy circles, regulators from Bank of Ghana, for example, have expressed concerns about such potential market misconduct. There are ongoing regulator and stakeholder discussions about eliminating emerging risks and recognizable fraud on this market and proving ultimate consumer confidence in mobile financial services. In Ghana, the MNOs and their commercial partners have been charged to build more risk and fraud-resilient financial infrastructures.[1] Our present study is designed to understand financial misconduct at vendor banking points (see, Figure 10 in Appendix G) and examine its potential efficiency and equity costs. We do this in a rural context where the business of M-Money could have larger impacts, *if* well designed.

II.1.2 Market Census

Detail vendor $\times$ customer data on M-Money is unavailable. So, between February-March 2019, we carried out a unique census of the market for M-Money in Eastern Ghana, spanning 9 districts. Districts are made up of sub-administrative units called "localities" or villages. Eastern Ghana was chosen for its two attractive features: (i) it covers an expansive number of villages, with potentially mobile banking sites, and (ii) our initial pilot works in other parts of this region suggest substantial levels of misconduct in this market. Our census exercise documents the universe of all vendor banking points (both formal and informal), and other surrounding households (within 5 houses radius around a given vendor) successfully across 130 localities. We focus on nearby households in order to maximize our chances of studying households that might make transactions with select vendors, while minimizing costs. We define a local market as the pair: vendor $\times$ the set of all nearby customers (see, Annan 2020 for details).

[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, 2015.**

II.1.3 Market Facts

Our baseline census solicited information from all market participants: both vendors and customers. We asked information on their basic demographics, poverty and assets, detail 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 market impacts of misconduct. Detail summaries, and other patterns about the market are available in Annan (2020) and upon request.

Table 15 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. 62% of the potential customers are females, and customers are more likely to be self-employed, married and older relative to vendors. Approximately and strikingly, half of the vendorship have received formal training about the market for M-Money before joining the business. The overwhelmingly majority (90% [SD=0.29]) of customers, their close family and friends networks have registered for a M-Money account (also called "wallet"), indicating that it is likely a popular financial technology.

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.[2] The average daily sales per vendor is about GHS2,260 [US$442]. 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: 95% of customers are M-Money users, 80% are

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

past formal bank users, while just 9% are post-office users. This can be explained by the convenient access 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.

II.1.4 Motivating Features: Asymmetric Information, Reputation, Perceptions about Misconduct

The presence of asymmetric information and regard for reputation are key ingredients of our study: information programs, their effects and interpretations. In the Theory section, we show evidence that our (baseline) setting reflects a market environment where (i) consumers are objectively uninformed (less sophisticated), (iii) consumers misperceive the level of misconduct, and (iii) vendors value their reputation as there is a positive return to good reputation but difficult to establish such reputation (i.e., with scope to build reputational capital).

II.2 Intervention and Timetable

We evaluate the impacts of different information sets that reduce market misconduct on both customers and vendors. As we discuss later in the Theory section, the provision of relevant market information about vendor misconduct to (potentially uninformed) consumers raises vendor concerns for reputation as customers are likely able to infer (ir)responsible vendors and then assign reputational payoffs to the vendors. *If* vendors care about such (negative or positive) perceptions, then misconduct will fall, which has market-wide implications for the outcomes of our study. This provides theoretical basis to fix our ideas and motivate the information programs.

All local markets (vendor $\times$ customers) receive a physical research visit, and markets assigned to treatment receive additional information about misconduct. For all markets, we show subjects the reconstructed market rosters, ask them to indicate where their last financial transactions were conducted, and provide contact information of our research team for further assistance. Markets assigned to treatments additionally receive either of the following:

- Treatment program I: Price Transparency – Addresses the question of "what to ask vendors while at banking points". It informs and educates consumers about the true tariffs for common local transactions and improves consumer sophistication about detecting misconduct.
- Treatment program II: Monitor and Report – Addresses the question of "how to report transactional glitches or misconduct". It informs customers while raising the potential cost of misconduct to vendors *if* caught.
- Treatment program III: Price Transparency+Monitor and Report – A joint program that tests the interaction of programs I and II. (see, Exhibits in Appendix F for the specific information sets).
- Control program: no additional information.

To ensure meaningful treatment effects, we visit the assigned local markets 3 consecutive times over a 3 months period (once per month) to first deliver and then repeat the information programs to subjects. Visits are concluded with the subjects summarizing the information they received and keeping hard copies of the treatment program. More uniquely, we ensured that vendors are equally aware of the interventions by communicating the same information set to the vendors right after seeding the information with the nearby households – yielding a two-sided information design.

To roughly gauge the likely significance of the information programs, the recipients were *ex-ante* asked to rate the usefulness of the information we provided for their financial decision-making (i.e., customers) and businesses (i.e., vendors) on a 5-point scale: `1 (Not useful), 2 (Quite useful), 3 (Useful), 4 (Very useful), 5 (Extremely useful)`. Overall, the median value = 3 (mean=3.38, [SD=0.82]), suggesting that subjects view our information interventions as useful, and thus likely to be *ex-post* effective.[3] Program I is a popular consumer protection policy instrument. By benchmarking this with programs II and III, we can evaluate program I's relative effectiveness in reducing market misconduct committed against consumers, and assess whether program I is compatible with other information programs or whether it only becomes effective when combined with an alternative that increases the cost of misconduct to firms.

Table 1: **Study timeline**

	DATE	ACTIVITY
Part 1	February 2017	**Pilots:** Misconduct– incidence, correlates
Part 2	Feb 15-Mar 20, 2019	**Baseline:** Market census– detail market records, demographics, main outcomes, misconduct beliefs
	Sep 01-Oct 15, 2019	**Audit study I:** Estimate misconduct, $\geq$ 1 in 5 transactions (22%)
Part 3	Oct 15-Dec 15, 2019	**Intervention:** Information assignment Control: no information Treatment I: price transparency Treatment II: monitor and report Treatment III: price transparency + monitor and report **Transaction networks data:** family vs friends vs strangers
Part 4	May 15-May 30, 2020	**Endline:** Phone survey + manual tracing supplement main outcomes, misconduct beliefs
	Aug 15-Sep 01, 2020	**Audit study II:** Re-estimate misconduct (12% of transactions)
	> Sep 15, 2020	**Administrative data (vendors):** transaction record volumes

[3]In practice, there were instances where the experimental subjects (specifically the customers) took "costly actions" to call our research team to discuss their M-Money 2-3 months after the provision of the information programs. This will suggest that subjects are willing to pay for our information programs, perhaps because they find the information credible. In addition, this will suggest that subjects' rating of the usefulness of the information provided is less likely affected by potential experimenter demand (pleasing) effects (see, de Quidt, Haushofer and Roth 2018).

II.3 Data Collected

We gather information from multiple rounds of data collection (i) combined listing and baseline market census (process discussed in "Market Census" above), (ii) baseline audit study (process discussed below), (iii) transaction networks data, (iv) 22-weeks follow-up (phone) market survey, 33-weeks audit study, and 35-weeks vendor administrative transaction data, which we call an endline.

II.3.1 Administrative Audit Data

To objectively measure "true" misconduct, we employ an audit study where auditors (experimental customers) were given cash to make actual transactions on M-Money, as credible data on misconduct is directly unavailable. The transactions span multiple transaction types which are common in the market, totaling 12: sending *versus* receiving transactions. Tariffs on transactions are ex-ante set by the providers and market regulators. 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 were trained to trained to follow the same approach on how to interact with the vendors, particularly use uniform language at visits to vendors (see, Annan 2020 for details). This approach has the strengths of measuring the "true" incidence of misconduct (unlike other survey-based measures of misconduct; DeLiema et al. 2018), while avoiding deception and its later effect on the market (unlike other standard audit studies; Kessler, Low and Sullivan 2019).

In our market setting, (and as in Egan, Matvos and Seru 2019b; Annan 2020), we define misconduct to entail transactions that are indicative of fraud or wrongdoing – i.e., overcharged and/or fake transactions with reference to the regulator and provider-approved tariff rates. Table 16 and Figure 5 in Appendix C show the baseline results. We estimate that 22% of transactions are overcharged (reflects the incidence of misconduct) and 3.3GHS (= 82% of the official tariffs) overpaid to the vendor as a result of misconduct (reflects the severity of misconduct).

II.3.2 Market Survey Data

We measure several repeated outcomes at different stages of the study. For customers, we restrict attention to 4 relevant outcomes: (i) adoption and usage of money services: ask whether households use the money services, and the transaction amount involved per week, (ii) savings on M-Money: ask whether households save on their money wallets, (iii) specific shock experiences and risk mitigation: ask whether customers experienced unexpected shocks that they could not financially remedy or pay for – providing an objective proxy for insurance (Dupas and Robinson [2013]; Breza and Chandrasekhar [2019]), and (iv) poverty. Since our study focuses on mobile banking in low-income and poor environments, we fielded questions 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". Details about this poverty scoring methodology can be found in (Schreiner 2015).

For vendors, we measure sales volume: solicit transaction records for their M-Money business and non M-Money services (*if* the vendor operates a joint venture). Together, we gather data from both sides of the market, which allows us to cross-validate accuracy of the records. For example, one will expect increases in households money transactions to (positively) correlate with increases in nearby vendor sales, all else equal.

II.3.3 Administrative Transaction Data

We supplement our data sets with administrative data on vendor transactions from an (informal) association that registers M-Money vendors across the region. We partnered with the association to solicit from its members their transaction records, 35-weeks after the interventions.

II.4 Treatment Assignment

We use a 2×2 factorial design, randomizing the total 130 "representative" markets (as defined below) into 4 experimental anti-misconduct programs: price transparency-only (31 markets

$\equiv$ 31 representative vendors $\times$ 272 nearby customers), monitor and report-only (32 markets $\equiv$ 32 representative vendors $\times$ 257 nearby customers), joint program (35 markets $\equiv$ 35 representative vendors $\times$ 276 nearby customers), and control program (32 markets $\equiv$ 32 representative vendors $\times$ 185 nearby customers). We stratified based on districts, and all misfits are resolved and randomly assigned.

II.5 Balance and Validity of Design

II.5.1 Balance I

We focus our study on selected "representative" markets (villages) drawn from a listing of the baseline market census. Each of the 130 localities has one or more vendor(s) (range=1-12, average=3.3) with their surrounding customers or households (range=5-47, average=20.8). To maximize statistical power, we "randomly" select one vendor and his/her nearby customers per locality for our study. We call this combination (representative vendor $\times$ nearby households) a "representative" market. Sample representativeness requires that being a "representative" market is independent of any relevant market-level statistics. To test that these samples are comparable to the market population, we run the regression

$$y_{mv} = \alpha + \beta S_{mv} + \epsilon_{mv}$$

on the baseline census data, where $S_{mv} = 1$ if market pair m from the pairs in village v is selected to be a representative in the *pre* intervention 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 10 and 11 report the results, where we find no difference across markets selected and those not-selected to be representative.

II.5.2 Balance II

We base our treatment analysis on a comparison of "representative" local markets ($m = v$ now) that received the information treatments with those that did not receive the treatments. Successful randomization of treatments, and thus identification requires that the assignments to treatments (i.e., price transparency-only, monitor and report-only, and joint information sets) are independent of any relevant household or market-level statistics. Similarly, to test that these markets are comparable, we run the regression

$$y_{iv} = \alpha + \beta \mathbf{I}_v + \epsilon_{iv}$$

on the baseline data, where $\mathbf{I}_v = 1$ if local market v in district d received an information treatment, 0 otherwise. We consider the various treatments separately and together (i.e., meta) for a number of different outcomes, and show that both sides of the market show no observable differences across the two groups. Tables 12 and 13 report the results, and provides strong evidence in favor of balance with no difference across subjects i (households or vendors) in assigned (treated) and non-assigned (control) markets.

II.5.3 Attrition

Our randomization is based on the selected "representative" markets that draws on the baseline market census. Table 14 displays the breakdown of response rates and attrition between baseline and endline. Here, attrition may be linked to subjects non-response, migration to outside the locality, and inability to reach the participants either because their phone numbers are inactive or out of network coverage area. To maximize response rates at endline, trained field officers conducted multiple phone calls (see, Figure 4) at different time horizons of the day, varying either weekdays or weekends, combined with manual contact tracing for subjects with inactive phone numbers. We record an overall attrition rate of 18%, which is low given that the business of M-Money is subject to high degree of migration and operator turnovers. Attrition looks non-differential. For our endline audit transactional exercises, 129 out of the 130 representative vendors were reached, implying an attrition rate of just 0.8%. In our empirical estimations, we evaluate and formally show robustness to attrition by treatment status.

III Experiment: Results

We present and discuss the treatment effects. Since all our treatments are about information provision, we first report the (combined) meta effect of information assignment, and then the separate effects for the different treatments.

III.1 Empirical Specifications

We estimate treatment effects using the model:

$$y_{ivd} = \beta \mathbf{I}_{vd} + \eta_d + \beta_0 y_{base,ivd} + \mathbf{X}'_{ivd}\xi + \epsilon_{ivd}$$

which links various endline outcome(s) y_{ivd} of subject (customer or vendor) i in locality (village) v in district d to the random treatment variable(s) $\mathbf{I}_{vd}$, district-level (stratification unit) dummies η_d, baseline outcomes $y_{base,ivd}$ and additional vector of controls $\mathbf{X}_{ivd}$. We include baseline outcomes to primarily increase precision but these control for potential confounds (*if* any). For the meta effects, $\mathbf{I}_{vd}$ is a 0-1 indicator for whether a locality received any of the information programs, and thus β captures the (meta) treatment effect. For the separate effects, $\mathbf{I}_{vd}$ is a 0-1 indicator for whether a locality received a specific information program. We denote by β_1, β_2 and δ the separate treatment effects for price transparency-only, monitor and report-only, and joint information sets, respectively (i.e., $\beta = (\beta_1, \beta_2, \delta)'$).

For inference and robustness, we report various standard errors including, the wild bootstrap cluster-t and randomization inference both clustered at the (village) market level. To address the potential issue of multiple testing, we adjust p-values for multiple testing across family of outcomes following the procedure presented in List, Shaikh and Xu (2019). To evaluate and show robustness for "potential" attrition bias, we report Lee (2009) attrition bounds (trimming based on observed attrition rates; see, Table 14), Imbens and Manski (2004) confidence sets, and Behaghel et al. (2015) attrition bounds (trimming based on the number of times subjects were called before answering the phone survey; see, Figure 4). In alternative models, we choose $\mathbf{X}_{ivd}$ using post-double-selection LASSO (for good estimation performance, in addition to minimizing researcher degrees of freedom and the possibility for p-hacking; Belloni et al. [2014]). We will sometimes discuss effects that contain useful economic information (i.e., looking at effect sign and effect size)–whether significant or not (Abadie 2020).

III.2 Graphical Evidence of Treatment Effects

We begin with a graphical illustration of the treatment effects. Figure 1 plots the empirical cumulative distributions of endline log(transaction amounts per week) by treatment status. The effects are displayed for the various treatments together (meta) and separately (in keeping with the approach of reporting the meta versus separate information effects). There is strong visual evidence of positive effects of the information programs on customers transactional outcomes. This implies increased uptake of the M-Money financial services as a result of the information program. What is more striking is that the effects does not seem to be driven by specific parts of the distribution. A Kolmogorov–Smirnov (KS) test for equality of distributions rejects the null that the distributional pairs are equal in all cases (p-values<0.080) except for the price transparency-only program (p-value=0.288). Thus, there is a considerable difference between the distribution of treated versus control local markets as we reject the null hypothesis of no distributional effects. We proceed to quantify the impacts for the various economic outcomes. Our estimates are robust to alternative controls, inference procedures, and adjustment for attrition.

III.3 Information Assignment – Meta Estimates

III.3.1 Effects on M-Money usage and Savings

Tables 2 and 3 report the estimated meta effects on usage of services and savings, respectively. There is increased transaction amount per week (see, Table 2), with a semi-elasticity of 0.402 (=+11.2% of control mean, p-value=0.048). In Appendix Table 18, we report the effects on the probability of using financial services, showing increased transaction likelihood of usage per week (7.1 pp =+9.8% of control mean, p-value=0.049). For savings, there is evidence of increased savings rate (see, Table 3) by 7.6 pp (=+12.6% of control mean, p-value=0.099).

III.3.2 Effects on Mitigation of shocks: revenue, health, and expenditure

Did customers' (or households) increase their transactional services and savings in meaningful enough levels that they are better able to mitigate unexpected shocks? Table 4 shows the estimated meta effects on customers experiences to unmitigated shocks. We report this for general shocks (any experience), and then individually for shocks related to household revenue, health, and household expenditures, respectively.

There is reduced instance(s) of general unexpected shocks that consumers cannot financially remedy or pay for (i.e., when resource limits bind) (-6.8 pp =-7.6% of control mean, p-value=0.068). This effect is mainly driven by household expenditures, which has the largest significant reduction of 10.7 pp. However, both the health and revenue sources are equally meaningful looking at their effect sizes (7.2 pp and 5.6 pp, respectively). These estimates provide a large and objective proxy for resilience and insurance value of reducing market misconduct to consumers.

III.3.3 Effects on Welfare

In Table 6, we test whether the information program impacted poverty (Schneider 2005)–which is an objective measure of consumer welfare. We do not find evidence for an impact on overall poverty levels. This is less surprising given that poverty is a structural and composite outcome, which can take a while to see effects. We hypothesize, however, that increased adoption of the financial services, savings and continued resilience of households to unexpected shocks (reported earlier), might translate into longer-term changes in poverty. Moreover, there could be significant distributional impacts on poverty. As we will show, the interventions meaningfully reduced the poverty levels for females relative to male customers.

III.4 Information Sets –What's necessary, What's sufficient?

We now report the separate impacts by the different information programs.

III.4.1 Effects on M-Money usage and Savings

Table 5 shows the estimated effects of the various information sets for the uptake of services and savings. For uptake of services, the effects are positively much larger for the joint program (semi-elasticity of 0.506 =+14.1% of control mean, p-value=0.035), compared with the other individual information sets. The results are similar for savings behavior at vendor points. Customers are significantly more likely to save on M-Money with much larger impacts for the joint program (semi-elasticity of 0.123 = +20.2% of control mean, p-value=0.024), compared with the other individual information sets. A Wald test rejects the null that the savings effect from the joint program is equal to effect from the monitor and report-only information set (p-value=0.066)

We combine all the usage and savings outcomes (via PCA) (see, column 5 of Table 5), finding that the effects are consistently larger for the joint program. This is followed by the monitor and report-only, and then the price transparency-only information sets. These results indicate that the monitor and report-only and price transparency-only programs are (informationally) complements, and that the price-transparency alone (a popular consumer protection instrument) may not be sufficient except when combined with random information assignment about monitoring and reporting.

III.4.2 Effects on Mitigation of shocks and Welfare

The estimated impacts for the various information sets on both shock mitigation and welfare are reported in Table 6. For shock mitigation, the joint information program show significantly (negative) larger impacts, compared to the alternative individual information counterparts. As in the meta estimate, this effect is mainly driven by mitigation of unexpected shocks related to household expenses. Effects from the monitor and report-only program are rather smaller and insignificant. For poverty, we also do not find evidence for an impact on overall poverty levels across the various programs. These results agree with our earlier findings that the two individual information sets are (informationally) complements and that the impact on poverty as a structural or composite outcome may distributional.

III.5 Closing the gap: Gender differences in Treatment Effects

Did the information intervention benefit females more? Recent research has shown that female customers suffer more misconduct relative to male customers (Annan 2020), and that M-Money has the potential to lift people out of poverty, particularly women, in developing countries (Suri and Jack 2016). *If* our intervention reduced misconduct, then one might expect higher usage of M-Money services for females. In response, the poverty rates should be lower for female consumers relative to their male counterparts. To evaluate such gender-differentiated effects of our treatments, we re-estimate the baseline model that includes an interaction for the customers gender.

Table 7 reports the estimated gendered effects. We observe interesting impacts by gender. While we find no overall impact on poverty (shown earlier), the interventions lowered poverty rate for females relative to males, significantly for the "monitor and report" information assignment. This effect is significantly different from the price transparency-only program (Wald test, p-value=0.038), but insignificantly different from the joint program's effect (Wald test, p-value=0.338). The meta estimate of the information treatment on gender (see, panel B of Table 7) is not significant yet contains useful economic information (Abadie 2020). While imprecisely estimated, the sign (negative) and size (-3.634) of the gender-differentiated effect implies meaningful economic effects; for example, when compared with the uninteracted effect of 3.282 and a control mean of 10.186. Similarly, the joint program in Table 7 shows large gender-differentiated economic effects, which arguably explains why the (significant) monitor and report-only assignment is not different from the joint program's effect.

To develop more intuition and understand the potential sources of these gender effects, we use the baseline data to verify that *pre*-intervention, there were meaningful gender gaps whereby female customers were transacting GHS134 less per week, 12.5 pp less likely to save, and 0.4 percent more likely to be poorer (all these differences were significant at the

1% level; see, panel A of Table 19). Such pre-existing gender gaps in usage and savings were eliminated following the interventions (signs reversed and no longer significant). This translated to reduced poverty rates for the females relative to male customers, consistent with our baseline hypothesis.

III.6 Treatment effects on vendor transactions

Did market vendors' experience an increase in sales? *If* the consumers records are accurate, and hence the estimated treatment effects, then one might expect direct increases in vendors transactional volumes (all else equal). We test for this and in doing so, evaluate the impact of the information intervention on vendors market activity.

Table 8 reports the estimated impacts on vendors. As expected, the information programs significantly increased vendors sales volume (+GHS557 per day = +46.5% of control mean) overall. The estimated +GHS557 per day seems reasonable: it translates to about GHS3,899 per week (for vendors). If we divide this weekly estimate by +GHS95 (as customers transact about +GHS95 per week in response to the information intervention; see, Table 17), we get approximately +41 more customers *per week* at a representative vendor point. This may reflect both repeat and/ or distinct customers. In additional tests (not reported), we find extended large positive impacts on non M-Money transactions, suggesting positive spillover effects of the information program on overall local financial activities.

IV Framework: Interpreting the Results

We present a framework to guide the interpretation of our results. We seek to understand what happens when we give relevant market misconduct information to both a (potentially unethical and informed) vendor and a (potentially uninformed) consumer in a local finance context. One could tell several stories about how the information intervention might act to affect misconduct, and thus the market outcomes. Our underlying hypothesis, however, is that vendors expect that they are more likely to be perceived (by potential customers) as *irresponsible* if they commit (or signal as) misconduct in our experiment. Such negative perceptions can affect vendors' sales in other joint lines of business, customer referrals, including other future market and social relations – yielding a misconduct signaling vs. reputation-type interpretation. We do not claim this is the only important interpretation.

Our goal is not to develop a general theory of either misconduct (e.g., Banerjee et al. 2012 for corruption) or signaling (e.g., Spence 1973). We rather provide a parsimonious model that embeds misconduct and signaling to deliver highly stylized predictions which guide the interpretation of our results. We turn first to relevant features of our setting to motivate the modeling framework and subsequent interpretation.

IV.1 Baseline Setting

We document relevant features of our empirical setting by providing 3 pieces of descriptive evidence: the presence of asymmetric (imperfect) information about the true transactional prices between vendors and customers, difficulty of vendors in establishing market reputation – amplified by the limited trust of customers in transacting, and misperceptions about misconduct – making it difficult for customers to infer otherwise (ir)responsible vendors.

IV.1.1 Feature 1: Asymmetric (Imperfect) Information

`Customers are less knowledgeable about true prices relative to market vendors (at baseline)?`

We draw on data from the baseline market census to examine if vendors have superior knowledge about the true transactional prices compared to customers. In a series of tests, both vendors and customers were asked to indicate the true charges for two randomly chosen transactions of sizes GHS200 (small to medium) and GHS1200 (large). For the vendors, we were careful to inform them at the beginning that we were not there to perform any actual transactions, but to rather assess their overall knowledge about the market (to alleviate any potential incentives for misconduct). Knowledge tests were taken towards the end of the surveys for both subjects. With reference to the official charges, this provides us an estimate of their knowledge about the true charges, specifically the % of subjects whose answers were correct across the markets.

Results are displayed in Figure 6, showing strong evidence of asymmetric information: vendors have superior knowledge about the true transactional charges relative to customers. While market vendors are relatively more knowledgeable, their knowledge is also imperfect. This noise can also limit the ability of vendors to build reputation as it exacerbates the incidence of misconduct. These results are expected because (unlike customers) vendors receive formal training about the market for M-Money before they start their businesses. Although a universal requirement, approximately and strikingly, only 50% of the vendors indicated they received formal training at the baseline (see, Table 15).

IV.1.2 Feature 2: Reputation

`I. Vendors: importance of good market reputation to vendors?`

We asked a random sample of vendors in the control group of our experiment *post*-endline about how important is it to show a high degree of good market reputation (or image and responsibility) to potential customers through their market transactions. As shown in Figure 7, the vast majority of vendors (81% [SD=0.391]) consider good market reputation or image as important. This descriptive suggests that the vendors value their reputation as there is likely a positive return or reward to good market reputation.

`II. Customers: (mis)trust for carrying out m-money transactions (at baseline)?`

Our baseline census solicited information about customers' level of trust in carrying out their transactions in the market. Figure 8 reports the results, suggesting limited level of trust. About 62% [SD=0.48, n=1275] of the customers indicate distrust in transacting at vendor points, while the rest (i.e., 38% [SD=0.48, n=779]) have trust making their money transactions. This suggests that vendors have low reputation in the market, perhaps because either (i) some vendors find it difficult to establish such reputation, or (ii) some customers are unable to infer the responsible vendors, both consistent with our earlier evidence of imperfect information about transactional tariffs.

IV.1.3 Feature 3: Perceptions about Misconduct

`Misperceived beliefs about misconduct (at baseline)?`

Figure 9 compares the true versus subjective beliefs of misconduct. Our actual audit transactions provide an objective (true) misconduct incidence of 22% [SD=0.41, n=663] at vendor points. We denote this by $(1-\pi)$, implying that π is the % of (non-misconduct) transactions *not*-overcharged. Details about the objective estimates of misconduct across a range of market audit transactions are illustrated in Table 16 and Figure 5, as noted. Next, we also asked customers views, at baseline, about the incidence of misconduct, yielding an overall subjective incidence of 59% [SD=0.49, n=1921] (denote that by $(1-\hat{\pi})$; implying a subjective incidence of non-misconduct $\hat{\pi}$=41%). Of course, the subjective belief estimate about non-misconduct $\hat{\pi}$ could be much higher, depending on how it is elicited. For our analysis, we thus assume consumers (mis)perceive the level of non-misconduct (and hence will allow for misperceived beliefs $\hat{\pi}$), and that the measured $\hat{\pi}$ is a (good) proxy for the relevant $\hat{\pi}$, which is lower. This assumption agrees with the observed departure of $\hat{\pi}$ from π and why misconduct is "prevalent" in the market at baseline.

Thus, this reflect an empirical setting where (i) consumers are (objectively) less sophisticated (uninformed), (ii) market vendors value their reputation in the market but such reputation is difficult to establish, suggesting that there is a positive return for good market reputation (e.g., extended sales, borrowing, referrals) if viewed by customers as responsible, and (iii) at baseline, consumers (under) misperceive the level of non-misconduct. Our setup and information program work to reduce vendors misconduct and enhance consumers (subjective) belief about the level of non-misconduct. Moreover, consumers might transact more *if* misconduct (equivalently, the marginal cost of transactions) is low.

IV.2 Model: Misconduct and Signaling

IV.2.1 Environment

Assume a continuum of local markets, defined by the pair: {representative vendor i, potential customer(s) j}. This is akin to our experiment's design, whereby we construct a local market using a randomly selected representative vendor and the nearby households as customers per locality to maximize statistical power. In each locality, the other vendors and customers have no designated role; our model will inherit the same design.

The vendors are of two mutually exclusive types $t_i \in \{G, B\}$, privately known by them. G types are pro-ethical (anti misconduct), while B types are anti-ethical (pro misconduct) and participate in signaling game with potential customers. Denote by π the % of vendors that are pro-ethical and thus designated as responsible, so $\Pr(t_i = G) = \pi$. We allow customers to hold imperfect belief about the % of vendors that are pro-ethical, which we denote by $\hat{\pi}$. $\hat{\pi}$ is assumed to be common knowledge to avoid instances of higher-order beliefs. The customers j assign a reputation payoff to i based on i's transactional choices: $s_i = 1$ (for $s = G$) versus $s_i = 0$ (for $s \neq G$), where $s_i = 1$ means not overcharging a transaction. That is, each vendor i decides whether to take a potentially costly "ethical" action ($s_i = 1$) at cost c_{t_i} or not ($s_i = 0$) at no cost. This action is observed by j. We assume $c_G < c_B$: this is a natural restriction because B types hold anti-ethical convictions, value misconduct more and thus engaging in a non-misconduct action should be more costly to them, relative to the G types. This simplifies the analysis, and as in Spence (1973), allows us to abstract away from the direct benefits of misconduct alone.

Let p_{ij} denote the probability that potential customer(s) j visit i's banking point and that a financial transaction take place. Similarly, let r_j be a simple indicator for whether or not j visited i. For $r_j = 1$, j transacts at i's banking point and offer a reputational-payoff that depends on j's inference of i's private type or responsibility. So, in every local market {representative vendor; nearby customer(s)}, potential customer(s) j visits each representative vendor i, independently, with probability p_{ij}. *If* the transaction takes place, customer(s) j offers vendor i a payoff which is proportional to her belief about i's responsibility, denoted by: $R_j[t = G|s_i, r_j]$. We call this reputational payoff as the vendor cares about $R_j[.|., .]$ that the customer computes (the posterior) and assigns, which takes place immediately (as in Shapiro 1983). As a practical foundation: *if* the customer thinks well of the vendor, the vendor will have access to valuable future opportunities e.g., extended sales, borrowing, referrals. By definition, this payoff depends on whether or not j visited i, and then i's transactional action. Consequently, the expected utility (payoff) of a vendor i with two-dimensional attribute (t_i, c_{t_i}) for choice s_i, observed by customer j is given by

$$\begin{aligned} U_i(s_i) = & - s_i c_{t_i} + \sum_j p_{ij}[\hat{\pi} R_j[t_i = G|s_i, r_j = 1] + (1 - \hat{\pi})(1 - R_j[t_i = G|s_i, r_j = 1])] \\ & + \sum_j (1 - p_{ij})[\hat{\pi} R_j[t_i = G|s_i, r_j = 0] + (1 - \hat{\pi})(1 - R_j[t_i = G|s_i, r_j = 0])] \end{aligned}$$

IV.2.2 Mapping Model to Experiment

Before analyzing the framework, it is useful to discuss how our model and analysis map to our experimental set-up. Market vendor(s) decide whether to commit misconduct ($s_i = 1$) or not ($s_i = 0$). This action sends a signal to the customer (uniformed vs informed) as to whether the vendor is a responsible type $t_i = G$ or not $t_i = B$. Based on whether the customer knows about the vendor's type via the signal received, the customer assigns a reputational payoff (via $R_j[.|., .]$) to the vendor. If the customer believes or perceives (via $\hat{\pi}$) that the vendor is responsible, then the vendor receives higher payoff (e.g., through repeated visits to transact) and vice versa.

Our goal is to compare two market information sets about misconduct: one "without" information and another "with" information assignment about misconduct. We model assignment of the anti-misconduct market information as either a shift in the distribution of $\hat{\pi}$ or $R_j[.|., .]$. As we show (and as implied by the model), the information assignment (i) increased customers beliefs about responsible vendors $\hat{\pi}$, (ii) caused customers to update their beliefs about non-misconduct (to assign $R_j[.|.,.]$), and (iii) caused vendors to update their beliefs about how informed consumers are. Together, these increased non-misconduct actions ($s_i = 1$) and improved market outcomes.

IV.2.3 Analysis

In the game, we are interested in Perfect Bayesian Equilibria. Let us define $DD_j = \sum_j p_{ij}$: this captures either the number of customers j in a given local market that visit vendor i (alternatively, number of j visits) or an intensity factor on j's belief about i's responsibility (In Appendix A, we provide detail foundations for DD_j). Denote $\hat{\pi}^* = \frac{c_G}{2\sum_j p_{ij}} + 1/2 = \frac{c_G}{2DD_j} + 1/2$ (assume $\hat{\pi}^* < 1$), and $r_j = 1$ if customer j visits a nearby vendor to make a successful transaction.

Proposition 1. Equilibrium: *Consider the model and stated assumptions. There is a Perfect Bayesian equilibrium (PBE) which is a cutoff such that*

(1) for the G types:

$$s_i = \begin{cases} 1 & if\ \hat{\pi} > \hat{\pi}^* \\ 0 & otherwise \end{cases}$$

(2) for the B types: $s_i = 0$ whether or not $\hat{\pi} > \hat{\pi}^*$. This PBE is supported by the following beliefs:

- $\Pr(t_i = G) = \Pr(t_i = G|r_j = 1) = \Pr(t_i = G|r_j = 0) = \hat{\pi}$
- $\Pr(t_i = G|s_i = 1, r_j = 1, \hat{\pi} > \hat{\pi}^*) = 1$ *and* $\Pr(t_i = G|s_i = 0, r_j = 1, \hat{\pi} > \hat{\pi}^*) = 0$
- $\Pr(t_i = G|s_i = 1, r_j = 1, \hat{\pi} < \hat{\pi}^*) = \underbrace{x \in (0,1)}$ *and* $\Pr(t_i = G|s_i = 0, r_j = 1, \hat{\pi} < \hat{\pi}^*) = \hat{\pi}$
  - *Analogously, for $t_i = B$*

Proof. See Appendix A. ■

When $r_j = 0$, there is no customer visit to the vendor and as such there is no updating:

the posterior equals the prior $\hat{\pi}$. The [maximal] extent of reputation gain is given by the difference: $\triangle R_j = R_j(s=1) - R_j(s=0)$ which depends on the vendor's actions about misconduct and the posterior payoff the customer computes and assigns.

Proposition 2. Information Intervention Effect: *(i) Changing subjective belief: $\hat{\pi}' > \hat{\pi}$ i.e., $\hat{\pi}' \in (\hat{\pi}, \hat{\pi}+\epsilon; \epsilon > 0)$. By shifting beliefs $\hat{\pi}' > \hat{\pi}$, it increases the number of $s = 1$. (ii) Changing the number of informed (sophisticated) customers. Denote by θ be the number of informed customers. By shifting θ: $\theta' > \theta$ i.e., $\theta' \in (\theta, \theta' + \epsilon; \epsilon > 0)$, it (weakly) increases the number of j visits to the G types i, $DD_{j \to i=G}$ (making the equilibrium $s = 1$ more likely). In contrast, increasing θ (strictly) decreases the number of j visits to the B types i, $DD_{j\to i=B}$ (unaffecting the equilibrium $s = 1$ since the B types never choose $s = 1$ in equilibrium regardless of the cutoff $\hat{\pi}^*$).*

Proof. See Appendix A. ■

IV.3 Effects – Subjective Beliefs, Reputation, and Belief updates

Subjective Beliefs: From the assumed (lower) $\hat{\pi}$, Proposition 2 indicates that an upward shift in $\hat{\pi}$ (as well as the number of informed customers θ) should increase the $\Pr(s_i = 1)$. Thus, a necessary requirement for our information program to *reduce* misconduct (with impacts on the allied market outcomes) is to check whether $\hat{\pi}$ increased. Did our information intervention actually increased $\hat{\pi}$? First, in Figure 2, we plot the distribution of $\hat{\pi}$ at endline – reflecting subjective beliefs about customers experiences of non-misconduct – by treatment status. These are displayed for the various treatments together (meta) and separately. Second, we estimate

$$\hat{\pi}_{jvd} = \gamma \mathbf{I}_{jvd} + \gamma_0 \hat{\pi}_{base,jvd} + \mathbf{X}'_{ivd}\xi + \zeta_{jvd}$$

controlling for consumer j baseline beliefs about non-misconduct. Table 20 reports the estimated effects of the information program on $\hat{\pi}$. There is strong evidence (both visual and formal) that the intervention shifted $\hat{\pi}$ upward in meaningful levels. We estimate a meta effect of +6.7 pp (+30% of control mean) increase in customers subjective beliefs about non-misconduct. The effect appears to be much larger for the joint program as expected.

Reputation: We measure reputation based on either how customers are able to infer vendor misconduct, or how vendors themselves are able to detect informed-customers. These are two major ingredients for reputation and its concerns for market vendors based on our simplified setup. Indeed, by providing market information about misconduct, it becomes more likely to detect misconduct and thus raising the importance of reputation.

Empirical Test I: Consumers updated their beliefs about vendors misconduct? We define this as the probability of a customer guessing (or inferring) correctly misconduct and vendor irresponsibility (or responsibility) given the information treatment $\Pr(t_i|\mathbf{I}_{jvd}) \equiv R_j[.|.,.]$. We compute this as an indicator that equals 1 whenever the customer guessed the presence of misconduct in the locality (or market) and the audit exercise objectively revealed misconduct in that locality. As we defined, this corresponds to the payoff that customers assign to vendors. Similarly, we estimate

$$\Pr(t_i|\mathbf{I}_{jvd}) = \gamma\mathbf{I}_{jvd} + \gamma_0 \Pr(t_i)_{base,jvd} + \mathbf{X}'_{ivd}\xi + \zeta_{jvd}$$

We estimate a meta effect of +8.8 pp (87% of control mean) increase in customers' ability to guess misconduct behavior, and the effects are economically larger for the joint information program (see, Table 21). These results (i) provide evidence of consumers' belief update - i.e., increased ability of customers to predict irresponsible vendors, and (ii) shows increased sophistication of consumers. *If* vendors recognize this, then they might update their beliefs about how informed or sophisticated consumers are by reducing misconduct. We test for this possibility below.

Empirical Test II: Vendors updated their beliefs about customers sophistication (i.e., θ)? This is defined as the reduction in vendors misconduct as a result of the anti-misconduct information programs. We estimate the baseline model:

$$Misconduct_{ivd} = \gamma \mathbf{I}_{ivd} + \eta_d + \gamma_0 Misconduct_{base,ivd} + \mathbf{X}'_{ivd}\xi + \zeta_{ivd}$$

Table 9 reports the effects, and shows that the intervention meaningfully reduced vendor misconduct (measured using actual audit transactions). We estimate a meta effect of -21 pp (-72%+ of control mean). The effects are economically much larger for the joint and monitor and report programs, however the differences across the programs are barely distinguishable statistically. In additional tests (not reported), we find significant spillover effects: non-treated vendors located in treated localities (or markets) reduced their misconduct. Results from follow-up surveys also provide corroborative evidence that the vendors' updated their beliefs about consumer sophistication. Overall, these results are strongly consistent with our proposed reputation-based interpretation: by providing symmetric information about misconduct to both parties (uninformed customers and informed vendors), it raises vendors concerns for market reputation. In response, vendors reduce their misconduct which has market-wide impacts.

V Cost-Effectiveness of the Information Program

How cost-effective is our information intervention? Does this compare well with other financial education interventions? When computing cost-effectiveness, we focus on usage of services-only measure for customers and sales volume-only measure for vendors. This is a very conservative approach in the sense that it does not consider the significant treatment effects on savings, risk mitigation, poverty outcomes, and other positive externalities of the information program. For example, we find additional improvements on non M-Money sales volume for vendors (not reported).

We estimate costs based on the # of trained field officers utilized (3 officers to minimize cross-officer heterogeneity), the # of times or rounds the experimental subjects were visited ($3\times$) to deliver the interventions over the period (October 15-December 15, 2020), transportation costs (GHS385 per officer $\times$ 3 officers $\times$ 3 rounds = GHS3,465), remuneration and allowance for officers (GHS1,200 per officer $\times$ 3 officers $\times$ 3 rounds = GHS10,800), and occasional accommodation for officers during field visits (GHS100 per officer $\times$ 3 officers $\times$ 3 rounds = GHS900). The total cost equals to GHS15,165. About 632 panel of treated customers were reached.[4] Similarly, about 97 panel of treated vendors were reached ($\frac{98+96+98}{3} = 97$), bringing the total # of subjects to 730. We then estimate $\frac{\text{GHS15,165}}{730} =$ GHS20.8 per subject, or US$3.8 per person at an exchange rate of US$1=GHS5.5. The opportunity cost of time-use for the subjects is very limited: it took roughly 7 minutes per visit to deliver the information intervention. When compared with the minimum wage in Ghana (GHS10.65 per day), the time-use and thus its cost on subjects is very negligible. Thus, the information sets cost approximately US$3.8 per subject.

Overall, our cost-effectiveness ratio is 1:5 – a per subject cost of US$3.8 for about +US$18.2 increase in the usage of financial services for customers (see, Table 17), with sizable implications for consumer welfare (risk mitigation, poverty; see, Table 4). For vendors, the treatment effect (+GHS591; see, Table 8) implies a ratio of 1:28 improvement in vendor outcomes. These rough cost-effectiveness estimates compare favorably with other financial information programs. For example, Frisancho (2018) reports a cost per pupil of US$4.80 and a US$1 increase in financial education program's expense for a 3.3 point improvement in financial literacy. In a recent meta-analyses about financial education interventions, Kaiser et al. (2020) reports a cost-effectiveness ratio of $60.40 per person for one-fifth of a standard deviation improvement in outcomes. Our findings suggest that providing "market-level" information that reduce misconduct in local market transactions could be a cost-effective way to improve market efficiency (and likely equity).

VI Conclusion

Misconduct in markets matters: in terms of both efficiency and equity. By providing in-

[4] $\frac{1}{3}\sum_{r=1}^{3}$ # of subjects reachable per round$_r$ = $\frac{629+617+642}{3} = 632$.

formation sets that deter and reduce market vendors misconduct – customers meaningfully increase their take-up of transactional services and savings behavior at vendor points, enabling customers to better mitigate unexpected shocks. Market vendors experience meaningful increases in their sales volume, suggesting improved market efficiency. Owing to prior gender gaps in uptake of transactional services and the fact that female customers disproportionately suffer more misconduct (Annan 2020), female customers experience reduction in poverty rates (similar to Suri and Jack [2016]), suggesting improved equity.

Reputation does matter for misconduct. In rural financial environments, where markets are subject to a high degree of information imperfections and asymmetries, the importance of reputation as a discipline device against market misconduct is limited. Misconduct may be high because vendors are unable to demonstrate the quality of their services and thus establish meaningful reputation. Reputation however becomes effective and disciplinary *if* there is a high probability of inferring misconduct (Burkhardt 2018) and vendors can easily demonstrate the quality of their market services. Such reputation-driven misconduct is illuminated drawing on features of our empirical setting and the provision of relevant market information that improves subjects' ability to make inferences about misconduct.

Our field experiment is carefully designed to: (i) reduce market misconduct using information programs about misconduct, (ii) quantify the impacts on important economic outcomes on both sides of the market: uptake of transactional services, savings, risk mitigation, poverty and sales, and (iii) show that these effects are driven by the importance of reputation. We do this by constructing a unique census of markets for mobile banking in rural Ghana, showing baseline evidence of significance misconduct, information imperfections and misbeliefs about vendor misconduct, and assigning these local markets to information programs about misconduct. Overall, our results emphasize the significance of (local) sanctions to support the growth of (rural) financial institutions (Karpoff 2012; Munshi 2014) and provide a proof-of-concept of a potentially significant source of local financial market friction, where market activities (i.e., adoption, transactions, savings, sales) are underprovided (Jensen and Miller

2018; Bai 2019) due to misconduct, with implications for market efficiency and equity in transactional markets. Reputation (or trust) in markets might be difficult to build and thus low, likely due to imperfect information.

References

[1] Abadie, Alberto. 2020. "Statistical Nonsignificance in Empirical Economics." American Economic Review: Insights, 2 (2): 193-208.

[2] Annan, Francis. 2020. "Gender and Financial Misconduct: A Field Experiment on Mobile Money." Georgia State University - Working Paper.

[3] Bai, Jie. 2019. "Melons as Lemons: Asymmetric Information, Consumer Learning and Quality Provision." Revise and Resubmit, Review of Economic Studies.

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

[5] Banerjee, Abhijit and Esther Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs. 15th Edition.

[6] Banerjee, A, Hanna R, and Mullainathan S. 2012. "Corruption". Ch. 27, pages 1109-1147 of: Handbook of Organizational Economics.

[7] Belloni Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." Review of Economic Studies 81(2): 608-650.

[8] Breza, E., and A. Chandrasekhar. 2019. "Social Networks, Reputation, and Commitment: Evidence from a Savings Monitors Experiment." Econometrica, 87(1): 175–216.

[9] Burkhardt, Kirsten. 2018. *Private Equity Firms: Their Role in the Formation of Strategic Alliances*. John Wiley & Sons, Inc.

[10] Coppejans, M., Gilleskie, D., Sieg, H. and K. 2007. Strumpf. "Consumer Demand under Price Uncertainty: Empirical Evidence from the Market for Cigarettes." Review of Economics and Statistics, 89 (3): 510-521.

[11] de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." American Economic Review, 108 (11): 3266-3302.

[12] Dupa, P., and J. Robinson. 2013. "Why Don't the Poor Save More? Evidence From Health Savings Experiments." American Economic Review, 103 (4), 1138–1171.

[13] Egan, Mark, Gregor Matvos and Amit Seru. 2019a. "The Market for Financial Adviser Misconduct." Journal of Political Economy, 127 (1): 233-295.

[14] Egan, Mark, Gregor Matvos and Amit Seru. 2019b. "When Harry fired Sally: The Double Standard in Punishing Misconduct." Revise and Resubmit, Journal of Political Economy.

[15] Frisancho, Veronica. 2018. "The Impact of School-Based Financial Education on High School Students and Their Teachers: Experimental Evidence from Peru". IDB Working Paper No. IDB-WP-871.

[16] Higgins, Sean. 2020. "Financial Technology Adoption." Northwestern University - Working paper.

[17] Jensen, Robert, and Nolan H. Miller. 2018. "Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India." American Economic Review, 108 (12): 3583-3625.

[18] Kaiser, Tim and Annamaria Lusardi, Lukas Menkhoff, and Carly Urban. 2020. "Financial Education Affects Financial Knowledge and Downstream Behaviors". Wharton Pension Research Council Working Paper No. 2020-07.

[19] Karpoff, M. Jonathan. 2012. "Does Reputation Work to Discipline Corporate Misconduct?" Ch. 18, pages 361-382 of: Oxford Handbook of Corporate Reputation. https://faculty.washington.edu/karpoff/Karpoff%20reputation%20survey%202012%20copy.pdf

[20] 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.

[21] List, John A., Azeem M. Shaikh and Yang Xu. 2016. "Multiple Hypothesis Testing in Experimental Economics". Experimental Economics, 22(4): 773-793.

[22] Munshi, Kevin 2014. "Community Networks and the Process of Development." Journal of Economic Perspectives, 28 (4), 49-76.

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

[24] Ortmann, Andreas and Ralph Hertwig. 2002. "The Costs of Deception: Evidence from Psychology." Experimental Economics, 5(): 111–131.

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

[26] Spence, M. 1973. "Job Market Signaling," The Quarterly Journal of Economics, 87 (3), 355-374.

[27] Shapiro, C. 1982. "Consumer information, Product Quality, and Seller reputation." The Bell Journal of Economics, 13(1), 20-35.

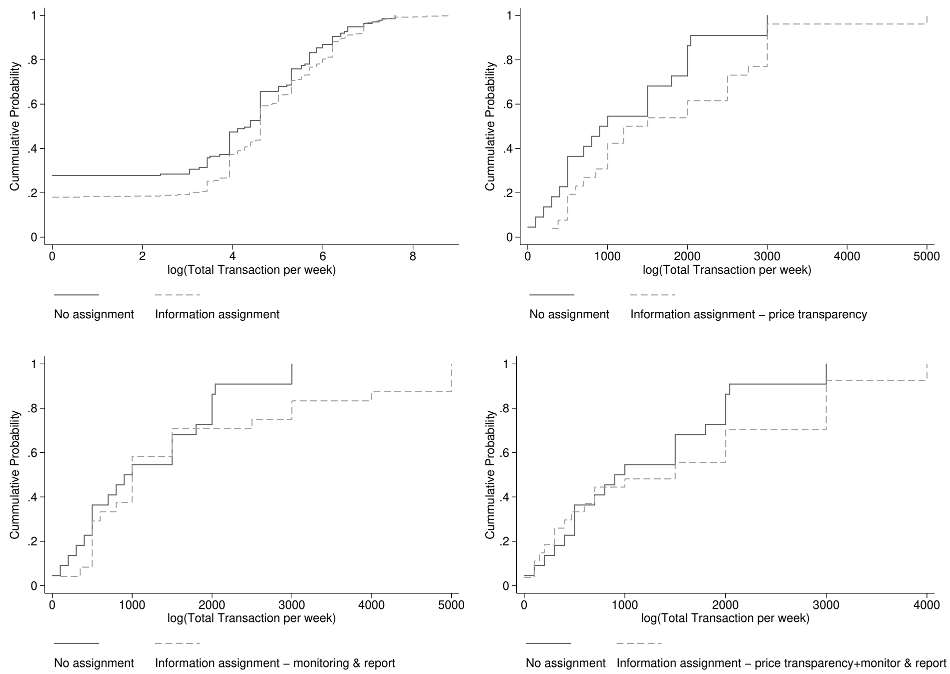
[28] Shapiro, C. 1983. "Premiums for High Quality Products as Returns to Reputations." Quarterly Journal of Economics, XCVIII, 658–679.

[29] Suri, Tavneet and William Jack. 2016. "The long-run Poverty and Gender Impacts of Mobile Money". Science, 354 (6317): 1288-1292.

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

Main Results for Text

Figure 1: **Distributions of log (total transactions) at endline by treatment status**



Note: Figure plots the distributions (CDFs) of log(Total Transactions per week) at endline for different experimental subsamples. From a Kolmogorov–Smirnov test for the equality of distributions, p-values equal 0.058, 0.288, 0.043 and 0.074, respectively.

Table 2: **Effect of treatment on usage**

Model: Linear
DV: Log Total Transaction per week

	(1)	(2)	(3)	(4)
Treatment: Information Assignment (β)	0.561	0.470	0.416	0.402
	(0.225)	(0.217)	(0.220)	(0.213)
	[0.189, 0.932]	[0.111, 0.828]	[0.052, 0.779]	[0.050, 0.755]
Inference Robustness (β)				
CI: Clustered S.E.	[0.059,1.062]	[0.096, 0.843]	[0.096, 0.843]	[0.082, 0.723]
CI: Wild Bootstrap	[0.191, 0.922]	[0.113, 0.821]	[0.113, 0.821]	[0.024, 0.789]
p-value: Permutation Test	0.015	0.032	0.041	0.048
p-value: L-S-X MHT Corr (2019)	0.012			
Market District F.E.		X	X	X
Baseline usage			X	X
Controls				X
Observations	763	763	723	723
R-squared	0.009	0.064	0.076	0.108
Mean of dependent variable (control)	3.583	3.583	3.583	3.583
Lee (2009) Attrition Bounds				
Lower Bound:	0.432			
	(0.271)			
	[-0.013, 0.878]			
Upper Bound:	0.805			
	(0.299)			
	[0.313, 1.297]			
Imbens and Manski (2004) CS	[0.076, 1.197]			
Behaghel et al. (2015) Attrition Bounds				
Lower Bound:	0.430			
	(0.226)			
	[0.059, 0.806]			
Upper Bound:	0.738			
	(0.225)			
	[0.366, 1.110]			

Note: Market district is the randomization strata. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) and confidence sets (CS) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level. L-S-X MHT Corr (2019) refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2019) for transactions outcomes family (services usage; savings). 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds. Results similar to post-double-selection LASSO estimates clustered at the market level.

Table 3: **Effect of treatment on savings**

Model: Linear
DV: 0-1 Indicator for whether consumer is "saving" on M-Money

	(1)	(2)	(3)	(4)
Treatment: Information Assignment (β)	0.089	0.078	0.080	0.076
	(0.045)	(0.045)	(0.046)	(0.045)
	[0.013, 0.164]	[-0.003, 0.147]	[0.004, 0.157]	[0.001, 0.151]
Inference Robustness (β)				
CI: Clustered S.E.	[-0.001, 0.178]	[-0.002, 0.146]	[0.007, 0.153]	[0.004, 0.148]
CI: Wild Bootstrap	[0.013, 0.165]	[-0.005, 0.149]	[0.004, 0.156]	[0.004, 0.149]
p-value: Permutation Test	0.059	0.108	0.080	0.099
p-value: L-S-X MHT Corr (2019)	0.048			
Market District F.E.		X	X	X
Baseline savings			X	X
Controls				X
Observations	763	763	689	689
R-squared	0.005	0.027	0.075	0.105
Mean of dependent variable (control)	0.605	0.605	0.605	0.605
Lee Attrition Bounds				
Lower Bound:	0.070			
	(0.050)			
	[-0.014, 0.152]			
Upper Bound:	0.125			
	(0.056)			
	[0.031, 0.218]			
Imbens and Manski (2004) CS	[0.001, 0.201]			
Behaghel et al. (2015) Attrition Bounds				
Lower Bound:	0.078			
	(0.046)			
	[0.003, 0.154]			
Upper Bound:	0.120			
	(0.045)			
	[0.045, 0.196]			

Note: Market district is the randomization strata. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) and confidence sets (CS) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level. L-S-X MHT Corr (2019) refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2019) for transactions outcomes family (services usage; savings). 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds. Results similar to post-double-selection LASSO estimates clustered at the market level.

Table 4: **Effect of treatment on shock mitigation and poverty**

Model: Linear

	u-Shocks Experience (1)	u-Shocks HH Revenue (2)	u-Shocks Health (3)	u-Shocks HH Expenditure (4)	Poverty Likelihood (5)
Treatment: Information	-0.068	-0.072	-0.056	-0.107	1.033
Assignment (β)	(0.030)	(0.040)	(0.044)	(0.044)	(1.254)
	[-0.117, -0.019]	[-0.140, -0.005]	[-.0130, 0.016]	[-0.180, -0.034]	[-1.033, 3.099]
Inference Robustness (β)					
CI: Clustered S.E.	[-0.128, -0.008]	[-0.159, 0.013]	[-0.163, 0.05]	[-0.206, -0.008]	[-1.306, 3.373]
CI: Wild Bootstrap	[-0.117, -0.020]	[-0.141, -0.007]	[-.1319, .018]	[-0.182, -0.033]	[-0.984, 3.107]
p-value: Permutation Test	0.068	0.176	0.332	0.091	0.451
p-value: L-S-X MHT Corr (2019)	0.027	0.057	0.601	0.161	0.140
Observations	763	763	763	763	763
R-squared	0.095	0.059	0.179	0.152	0.121
Mean of dependent variable (control)	0.890	0.773	0.525	0.416	10.18
Lee (2009) Attrition Bounds	[-0.089, -0.043]	[-0.103, -0.050	[-0.055, 0.003]	[-0.112, -0.053]	[-0.361, 3.286]
Imbens and Manski (2004) CS	[-0.134, 0.024]	[-0.164, 0.020]	[-0.128, 0.078]	[-0.190, 0.015]	[-2.761, 5.248]
Behaghel et al. (2015) Attrition Bounds	[-0.089, -0.045]	[-0.101, -0.058]	[-.058, -0.018]	[-0.099, -0.059]	[-0.178, 2.371]

Note: u denotes unmitigated. Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) and confidence sets (CS) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level. L-S-X MHT Corr (2019) refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2019) for welfare outcomes family (shocks mitigation; poverty). 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds. Results similar to post-double-selection LASSO estimates clustered at the market level.

Table 5: **Effect of information sets on usage and savings**

	Log Total Transaction per week	Total Transaction per week	Using M-Money	Saving on M-Money	PCA Index (1, 3, 4)
	(1) Linear-OLS	(2) Tobit	(3) Linear-OLS	(4) Linear-OLS	(5) Linear-OLS
Price Transparency (β_1)	0.280	39.684	0.059	0.064	0.088
	(0.247)	(54.369)	(0.044)	(0.053)	(0.110)
Robust S.E.	[-0.127, 0.687]	[-49.863, 129.231]	[-0.014, 0.133]	[-0.022, 0.152]	[-0.093, 0.270]
Clustered S.E.	[-0.103, 0.664]	[-38.782, 118.151]	[-0.011, 0.130]	[-0.071, 0.247]	[-0.069, 0.247]
Wild Bootstrap	[-0.124, 0.688]		[-0.014, 0.135]	[-0.021, 0.150]	[-0.097, 0.273]
p-value: Permutation Test	0.281	0.583	0.171	0.260	0.413
p-value: L-S-X MHT Corr (2019)	0.188		0.163	0.336	
Lee (2009) Attrition Bounds	<0.151, 0.767>		<0.051, 0.142>	<0.024, 0.122>	<0.060, 0.207>
Monitor and Report (β_2)	0.431	173.007	0.0705	0.036	0.188
	(0.253)	(83.049)	(0.044)	(0.054)	(0.110)
Robust S.E.	[0.014, 0.849]	[36.222, 309.792]	[-0.002, 0.143]	[-0.054, 0.126]	[0.007, 0.369]
Clustered S.E.	[0.031, 0.831]	[33.908, 312.106]	[-0.001, 0.142]	[-0.056, 0.128]	[0.026, 0.350]
Wild Bootstrap	[0.021, 0.842]		[-0.003, 0.143]	[-0.054, 0.125]	[0.001, 0.372]
p-value: Permutation Test	0.091	0.013	0.119	0.549	0.080
p-value: L-S-X MHT Corr (2019)	0.003		0.007	0.257	
Lee (2009) Attrition Bounds	<0.605, 0.790>		<0.106, 0.134>	<0.035, 0.072>	<0.262, 0.334>
Joint program: PT + MR (δ)	0.506	83.276	0.080	0.123	0.220
	(0.248)	(53.138)	(0.044)	(0.052)	(0.108)
Robust S.E.	[0.097, 0.915]	[-4.243, 170.797]	[0.008, 0.153]	[0.037, 0.208]	[0.042, 0.398]
Clustered S.E.	[0.129, 0.883]	[5.898, 160.655]	[0.012, 0.148]	[0.038, 0.207]	[0.067, 0.372]
Wild Bootstrap	[0.108, 0.907]		[0.007, 0.152]	[0.035, 0.211]	[0.036, 0.406]
p-value: Permutation Test	0.035	0.244	0.073	0.024	0.034
p-value: L-S-X MHT Corr (2019)	0.009		0.021	0.002	
Lee (2009) Attrition Bounds	<0.451, 0.877>		<0.096, 0.152>	<0.134, 0.191>	<0.198, 0.626>
Observations	723	723	723	689	689
Mean of dependent variable (control)	3.583	198.956	0.722	0.605	-0.201
p-value (test: $\beta_1 = \delta$)	0.298	0.336	0.583	0.203	0.175
p-value (test: $\beta_2 = \delta$)	0.739	0.204	0.786	0.066	0.745
p-value (test: $\beta_1 = \beta_2$)	0.502	0.077	0.780	0.562	0.315
p-value (test: $\beta_1 + \beta_2 = \delta$)	0.536	0.158	0.397	0.753	0.696

Note: Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level. L-S-X MHT Corr (2019) refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2019) for welfare outcomes family (shocks mitigation; poverty). 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds. Results similar to post-double-selection LASSO estimates clustered at the market level.

Table 6: **Effect of information sets on shock mitigation and poverty**

Model: Linear

	u-Shocks Experience (1)	u-Shocks HH Revenue (2)	u-Shocks Health (3)	u-Shocks HH Expenditure (4)	Poverty Likelihood (5)
Price Transparency (β_1)	-0.090	-0.110	-0.073	-0.128	1.680
	(0.036)	(0.047)	(0.052)	(0.051)	(1.509)
Robust S.E.	[-0.150, -0.029]	[-0.188, -0.031]	[-0.159, 0.012]	[-0.212, -0.044]	[-0.806, 4.167]
Clustered S.E.	[-0.159, -0.021]	[-0.214, -0.006]	[-0.194, 0.047]	[-0.244, -0.012]	[-1.077, 4.438]
Wild Bootstrap	[-0.151, -0.028]	[-0.188, -0.033]	[-0.161, 0.014]	[-0.214, -0.042]	[-0.712, 4.102]
p-value: Permutation Test	0.053	0.103	0.327	0.107	0.335
p-value: L-S-X MHT Corr (2019)	0.024	0.038	0.328	0.048	0.046
Lee (2009) Attrition Bounds	<-0.103, -0.004>	<-0.130, -0.031>	<-0.104, -0.005>	<-0.173, -0.074>	<-0.613, 4.974>
Monitor and Report (β_2)	-0.019	-0.001	-0.001	-0.041	1.439
	(0.036)	(0.049)	(0.052)	(0.049)	(1.552)
Robust S.E.	[-0.080, 0.041]	[-0.082, 0.079]	[-0.087, 0.084]	[-0.128, 0.045]	[-1.117, 3.997]
Clustered S.E.	[-0.088, 0.050]	[-0.105, 0.102]	[-0.126, 0.124]	[-0.168, 0.085]	[-1.231, 4.111]
Wild Bootstrap	[-0.081, 0.042]	[-0.080, 0.081]	[-0.086, 0.083]	[-0.132, 0.050]	[-1.202, 4.055]
p-value: Permutation Test	0.684	0.986	0.985	0.597	0.416
p-value: L-S-X MHT Corr (2019)	0.410	0.621	0.302	0.637	0.107
Lee (2009) Attrition Bounds	<-0.036, 0.0003>	<-0.032, 0.003>	<0.042, 0.079>	<0.006, 0.042>	<0.862, 3.716>
Joint program: PT + MR (δ)	-0.089	-0.096	-0.089	-0.143	0.022
	(0.036)	(0.048)	(0.051)	(0.049)	(1.456)
Robust S.E.	[-0.150, -0.029]	[-0.176, -0.016]	[-0.174, -0.005]	[-0.226, -0.061]	[-2.377, 2.421]
Clustered S.E.	[-0.167, -0.011]	[-0.195, 0.003]	[-0.207, 0.028]	[-0.250, -0.036]	[-2.895, 2.939]
Wild Bootstrap	[-0.150, -0.029]	[-0.176, -0.014]	[-0.176, -0.002]	[-0.229, -0.061]	[-2.492, 2.529]
p-value: Permutation Test	0.057	0.142	0.215	0.067	0.989
p-value: L-S-X MHT Corr (2019)	0.018	0.030	0.204	0.034	0.904
Lee (2009) Attrition Bounds	<-0.103, -0.029>	<-0.128, -0.054>	<-0.107, -0.034>	<-0.160, -0.086>	<-2.809, 2.336>
Observations	763	763	763	763	763
Mean of dependent variable (control)	0.890	0.773	0.525	0.416	10.18
p-value (test: $\beta_1 = \delta$)	0.983	0.751	0.714	0.718	0.235
p-value (test: $\beta_2 = \delta$)	0.052	0.034	0.050	0.021	0.326
p-value (test: $\beta_1 = \beta_2$)	0.057	0.015	0.123	0.059	0.870
p-value (test: $\beta_1 + \beta_2 = \delta$)	0.698	0.813	0.825	0.701	0.140

Note: u denotes unmitigated. Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level. 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds. L-S-X MHT Corr (2019) refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2019) for welfare outcomes family (shocks mitigation; poverty). Results similar to post-double-selection LASSO regression estimates clustered at the market level.

Table 7: **Gender differences in effects – usage, savings and poverty**

Model: Linear

	Log Total Transaction per week (1)	Using M-Money (2)	Saving on M-Money (3)	PCA Index (1-3) (4)	Poverty Likelihood (5)
Price Transparency (β_1)	-0.266	-0.045	-0.0003	-0.214	1.918
	(0.469)^c	(0.068)^c	(0.080)^c	(0.155)^c	(2.167)^c
	[-1.045, 0.511]^c	[-0.159, 0.068]^c	[-0.133, 0.132]^c	[-0.471, 0.042]^c	[-1.674, 5.511]^c
x Female (b_1)	0.847	0.161	0.102	0.495	**-0.571**
	(0.598)^c	(0.099)^c	(0.102)^c	(0.227)^c	(3.358)^c
	[-0.145, 1.840]^c	[-0.003, 0.327]^c	[-0.067, 0.272]^c	[0.118, 0.873]^c	[-6.138, 4.995]^c
Monitor and Report (β_2)	0.315	0.063	-0.037	0.069	5.719
	(.448)^c	(0.070)^c	(0.086)^c	(0.161)^c	(1.937)^c
	[-0.4280, 1.059]^c	[-0.0527, 0.180]^c	[-0.180, 0.106]^c	[-0.198, 0.337]^c	[2.508, 8.931]^c
x Female (b_2)	0.164	0.006	0.117	0.160	**-6.967**
	(0.625)^c	(0.104)^c	(0.106)^c	(0.240)^c	(2.945)^c
	[-0.872, 1.202]^c	[-0.166, 0.179]^c	[-0.059, 0.295]^c	[-0.239, 0.559]^c	[-11.850, -2.084]^c
Joint program: PT + MR (δ)	0.297	0.062	0.113	0.063	2.546
	(0.420)^c	(0.062)^c	(0.078)^c	(0.148)^c	(2.579)^c
	[-0.399, 0.993]^c	[-0.041, 0.166]^c	[-0.016, 0.242]^c	[-0.182, 0.310]^c	[-1.729, 6.822]^c
x Female (d)	0.323	0.026	0.014	0.226	**-4.200**
	(0.592)^c	(0.099)^c	(0.108)^c	(0.235)^c	(3.220)^c
	[-0.658, 1.304]^c	[-0.138, 0.191]^c	[-0.165, 0.194]^c	[-0.163, 0.615]^c	[-9.539, 1.138]^c
Female	-0.767	-0.107	-0.147	-0.395	4.047
	(0.480)^c	(0.081)^c	(0.081)^c	(0.185)^c	(2.195)^c
	[-1.564, 0.029]^c	[-0.243, 0.027]^c	[-0.281, -0.012]^c	[-0.702, -0.088]^c	[0.407, 7.687]^c
Observations	723	723	689	689	763
R-squared	0.114	0.112	0.076	0.114	0.131
Mean of dependent variable (control)	3.583	0.722	0.605	-0.198	10.186
p-value (test: $b_1 = d$)	0.281	**0.073**	0.122	0.151	0.284
p-value (test: $b_2 = d$)	0.754	0.799	0.268	0.740	0.338
p-value (test: $b_1 = b_2$)	0.196	**0.060**	0.868	**0.095**	**0.038**
p-value (test: $b_1 + b_2 = d$)	0.380	0.260	0.136	0.152	0.456

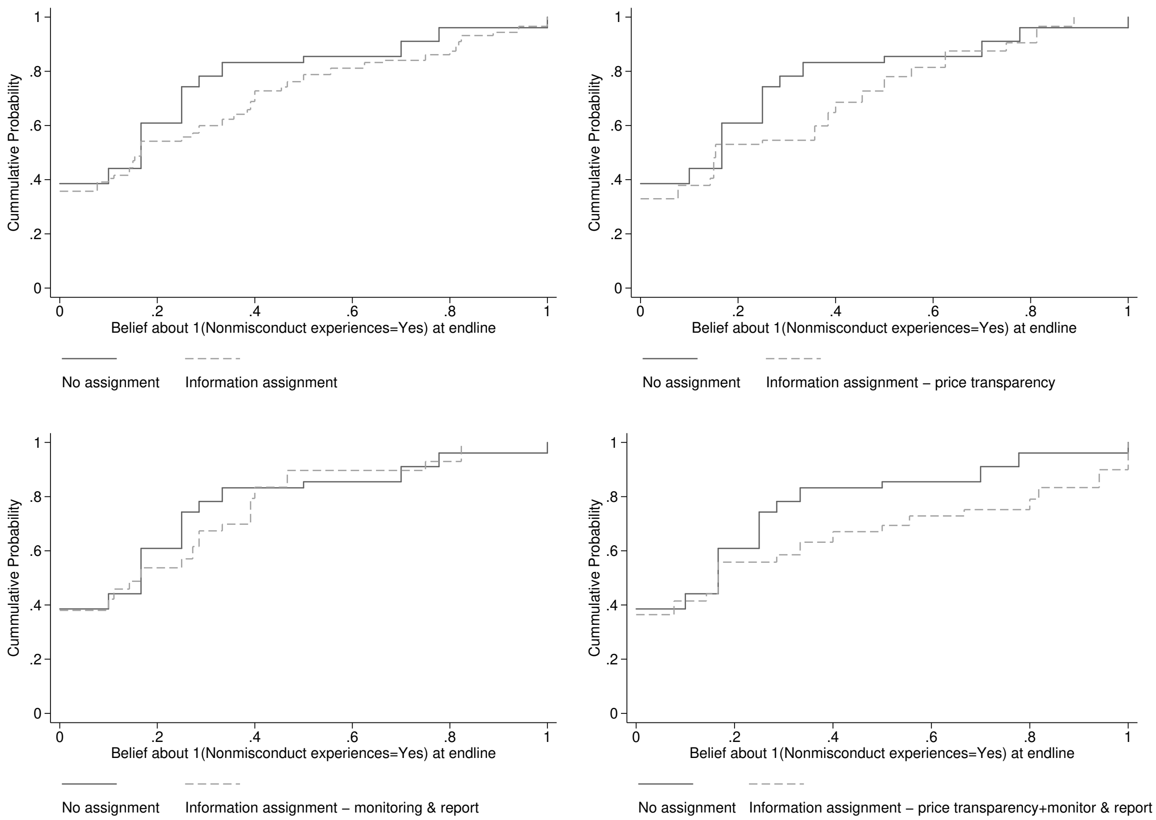
Note: Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. c denotes clustering. Clustered standard errors (at local market level) reported in parenthesis. 90% confidence intervals (CI) are reported in brackets. Results similar to post-double-selection LASSO regression estimates clustered at the market level.

Table 8: **Effect of treatment on vendors market sales**

DV: Total sales per day (GHS)	(1) Linear-OLS	(2) Tobit
PANEL A		
Treatment: Information	557.342	591.568
Assignment (β)	(278.916)	(274.918)
Robust S.E.	[93.386, 1021.298]	[44.862, 1138.273]
Clustered S.E.	[94.1867 102.498]	[44.862, 1138.273]
Wild Bootstrap	[92.150, 1026.000]	
Lee (2009) Attrition Bounds	<211.479, 622.647>	
Behaghel et al. (2015) Attrition Bounds	[432.877, 574.305]	
Observations	99	99
Mean of dependent variable (control)	1192.727	1192.727
PANEL B		
Price Transparency (β_1)	663.086	699.722
	(311.063)	(299.543)
Robust S.E.	[145.512, 1180.659]	[201.387, 1198.057]
Clustered S.E.	[146.549, 1179.623]	[201.387, 1198.057]
Wild Bootstrap	[169.900, 1169.000]	
Lee (2009) Attrition Bounds	<339.936, 702.517>	
Behaghel et al. (2015) Attrition Bounds	[443.286, 776.013]	
Monitor and Report (β_2)	569.105	648.863
	(438.705)	(439.512)
Robust S.E.	[-160.850, 1299.061]	[-82.330, 1380.058]
Clustered S.E.	[-159.388, 1297.599]	[-82.330, 1380.058]
Wild Bootstrap	[-177.400, 1290.000]	
Lee (2009) Attrition Bounds	<104.594, 714.902>	
Behaghel et al. (2015) Attrition Bounds	[157.272, 612.034]	
Joint program: PT + MR (δ)	421.780	422.605
	(310.394)	(296.952)
Robust S.E.	[-94.680, 938.241]	[-71.419, 916.630]
Clustered S.E.	[-93.646, 937.207]	[-71.419, 916.630]
Wild Bootstrap	[-195.700, 1361.000]	
Lee (2009) Attrition Bounds	<249.224, 435.588>	
Behaghel et al. (2015) Attrition Bounds	[261.620, 460.001]	
Observations	99	99
Mean of dependent variable (control)	1192.727	1192.727
p-value (test: $\beta_1 = \delta$)	0.408	0.321
p-value (test: $\beta_2 = \delta$)	0.707	0.562
p-value (test: $\beta_1 = \beta_2$)	0.808	0.894
p-value (test: $\beta_1 + \beta_2 = \delta$)	0.130	0.082

Note: Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) are reported in brackets. Reported confidence CIs for wild bootstrap derived from running 1000 replications. Differential attrition bounds are reported. 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds. Results similar to post-double-selection LASSO estimates clustered at the (local) market level.

Figure 2: **Distributions of customers belief about vendor responsibility (non-misconduct) at endline by treatment status**



Note: Figure plots the distributions (CDFs) of belief about (Non-misconduct experiences=Yes) at endline for different experimental subsamples. $\mathbf{1}(.)$ is a logical indicator that equals 1 if the argument in the parenthesis is true, 0 otherwise. In each market, we compute the share of experimental customers that indicate no experiences of misconduct (i.e., non-misconduct) at endline. From a Kolmogorov–Smirnov test for the equality of distributions, p-value $= 0.000$ for all cases.

Table 9: **Vendors belief update: effect of information sets on vendors misconduct**

	1(Misconduct=Yes)		Amount-Misconduct, GHS	
PANEL A				
Treatment: Information Assignment ($\gamma \equiv \beta$)	-0.231	-0.211	-0.675	-0.551
	$(0.055)^c$	$(0.086)^c$	$(0.185)^c$	$(0.255)^c$
	$[-0.324, -0.138]^c$	$[-0.354, -0.067]^c$	$[-0.984, -0.367]^c$	$[-0.975, -0.125]^c$
Baseline misconduct	X	X	X	X
Market District F.E.	X		X	
Market District × Transaction × Date F.E.		X		X
Observations	335	335	335	335
Mean of dependent variable (control)	0.294	0.294	0.778	0.778
Lee (2009) Attrition Bounds	<-0.174, -0.164>		<-0.484, -0.1435>	
Imbens and Manski (2004) CS	[-0.225, -0.094]		[-0.642, -0.085]	
PANEL B				
Price Transparency ($\gamma \equiv \beta_1$)	-0.177	-0.184	-0.550	-0.439
	$(0.065)^c$	$(0.094)^c$	$(0.199)^c$	$(0.276)^c$
	$[-0.285, -0.069]^c$	$[-0.342, -0.027]^c$	$[-1.881, -0.219]^c$	$[-0.898, 0.020]^c$
Monitor and Report ($\gamma \equiv \beta_2$)	-0.257	-0.217	-0.687	-0.574
	$(0.063)^c$	$(0.093)^c$	$(0.222)^c$	$(0.275)^c$
	$[-0.363\ 0.151]^c$	$[-0.373, -0.061]^c$	$[-1.057, -0.317]^c$	$[-1.032, -0.117]^c$
Joint program: PT + MR ($\gamma \equiv \delta$)	-0.233	-0.212	-0.718	-0.555
	$(0.064)^c$	$(0.089)^c$	$(0.198)^c$	$(0.279)^c$
	$[-0.340, -0.127]^c$	$[-0.360, -0.062]^c$	$[-1.048, -0.388]^c$	$[-1.019, -0.089]^c$
Baseline misconduct	X	X	X	X
Market District F.E.	X		X	
Market District × Transaction × Date F.E.		X		X
Observations	335	335	335	335
Mean of dependent variable (control)	0.294	0.294	0.778	0.778
p-value (test: $\beta_1 = \delta$)	0.327	0.670	0.280	0.553
p-value (test: $\beta_2 = \delta$)	0.660	0.921	0.860	0.923
p-value (test: $\beta_1 = \beta_2$)	0.**104**	0.563	0.347	0.411
p-value (test: $\beta_1 + \beta_2 = \delta$)	0.027	0.108	0.074	0.204

Note: **1**(.) is a logical indicator that equals 1 if the argument in the parenthesis is true, 0 otherwise. Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. c denotes clustering. Clustered standard errors (at local market level) reported in parenthesis. 90% confidence intervals (CI) and confidence sets (CS) are reported in brackets. Panel A reports meta estimate of treatment effects, while panel B shows effects separately for each information program. Results similar to post-double-selection LASSO estimates clustered at the (local) market level.

Supplementary Appendix
(Not For Publication)

A Proofs

A.1 Proof of Proposition 1

Proof. Consider $t_i = G : s = 1$ IFF

$$-c_G + \sum_j p_{ij}[\hat{\pi} R_j[t_i = G|s_i = 1, r_j = 1] + (1-\hat{\pi})(1 - R_j[t_i = G|s_i = 1, r_j = 1])]$$
$$+ \sum_j (1-p_{ij})[\hat{\pi} R_j[t_i = G|s_i = 1, r_j = 0] + (1-\hat{\pi})(1 - R_j[t_i = G|s_i = 1, r_j = 0])]$$
$$>$$
$$\sum_j p_{ij}[\hat{\pi} R_j[t_i = G|s_i = 0, r_j = 1] + (1-\hat{\pi})(1 - R_j[t_i = G|s_i = 0, r_j = 1])]$$
$$+ \sum_j (1-p_{ij})[\hat{\pi} R_j[t_i = G|s_i = 0, r_j = 0] + (1-\hat{\pi})(1 - R_j[t_i = G|s_i = 0, r_j = 0])]$$

No updating of vendor responsibility occurs if customer j does not visit vendor i, so $t_i = G :$ $s = 1$ IFF

$$-c_G + \sum_j p_{ij}[\hat{\pi} R_j[t_i = G|s_i = 1, r_j = 1] + (1-\hat{\pi})(1 - R_j[t_i = G|s_i = 1, r_j = 1])]$$
$$>$$
$$\sum_j p_{ij}[\hat{\pi} R_j[t_i = G|s_i = 0, r_j = 1] + (1-\hat{\pi})(1 - R_j[t_i = G|s_i = 0, r_j = 1])]$$

Since common beliefs assumed (everything is common knowledge except for vendor's private type), $t_i = G : s = 1$ IFF (rewrite by dropping subscript j)

$$-c_G + \sum_j p_{ij}[\hat{\pi} \underbrace{R[t_i = G|s_i = 1, r_j = 1])}_{M} + (1-\hat{\pi})\underbrace{(1 - R[t_i = G|s_i = 1, r_j = 1])}_{1-M}]$$
$$>$$
$$\sum_j p_{ij}[\hat{\pi} \underbrace{R[t_i = G|s_i = 0, r_j = 1])}_{M'} + (1-\hat{\pi})\underbrace{(1 - R[t_i = G|s_i = 0, r_j = 1])}_{1-M'}]$$

We get

$$\hat{\pi} > \frac{c_G + (M - M')\sum_j p_{ij}}{2(M - M')\sum_j p_{ij}} = \frac{c_G + DD_j \Delta R^{[G,1-G,0]}}{2DD_j \Delta R^{[G,1-G,0]}} = \frac{c_G}{2DD_j \Delta R^{[G,1-G,0]}} + 1/2$$

Now, consider $t_i = B : s = 1$ IFF (applying similar steps)

$$-c_B + \sum_j p_{ij}[\hat{\pi}\underbrace{R[t_i = B|s_i = 1, r_j = 1])}_{N} + (1-\hat{\pi})\underbrace{(1 - R[t_i = B|s_i = 1, r_j = 1])}_{1-N})]$$

$$>$$

$$\sum_j p_{ij}[\hat{\pi}\underbrace{R[t_i = B|s_i = 0, r_j = 1])}_{N'} + (1-\hat{\pi})\underbrace{(1 - R[t_i = B|s_i = 1, r_j = 1])}_{1-N'})]$$

We get

$$\hat{\pi} > \frac{c_B + (N - N')\sum_j p_{ij}}{2(N - N')\sum_j p_{ij}} = \frac{c_B + DD_j \Delta R^{[B,1-B,0]}}{2DD_j \Delta R^{[B,1-B,0]}} = \frac{c_B}{2DD_j \Delta R^{[B,1-B,0]}} + 1/2$$

Together, we get that $\forall t_i \in \{G, B\}$, $s_i = 1$ IFF

$$\hat{\pi} > \frac{c_{t_i}}{2DD_j \Delta R^{[t_i,1-t_i,0]}} + 1/2$$

In this PBE:

If $\hat{\pi} > \hat{\pi}^*$, then $R(s = 1) = \Pr(t_i = G|s_i = 1, r_j = 1) = 1$ and $R(s = 0) = \Pr(t_i = G|s_i = 0, r_j = 1) = 0$. Thus, $\Delta R^{[G,1-G,0]} = 1$ and $\Delta R^{[B,1-B,0]} = -1$. For $t_i = G$, $\hat{\pi} > \frac{c_G}{2DD_j(1-0)} + 1/2 \geq \hat{\pi}^* = \frac{c_G}{2DD_j}$. For $t_i = B$, $\hat{\pi} > \frac{c_B}{2DD_j(0-1)} + 1/2 < \hat{\pi}^*$. Thus,

$$\frac{c_B}{2DD_j(0-1)} + 1/2 < \hat{\pi}^* \leq \frac{c_G}{2DD_j(1-0)} + 1/2$$

and $t_i = B$ does not find it worthwhile to seek for a reputation gain $\Delta R^{[B,1-B,0]}$ by taking a pro-ethical action because the cost c_B is too high. But $t_i = G$ does.

If $\hat{\pi} < \hat{\pi}^*$, then $R(s = 1) = \Pr(t_i = G|s_i = 1, r_j = 1) = x \in (0, 1)$ and $R(s = 0) = \Pr(t_i = G|s_i = 0, r_j = 1) = \pi$ and

$$\hat{\pi} < \frac{c_B}{2DD_j(0-1)} + 1/2 < \hat{\pi}^* = \frac{c_G}{2DD_j(1-0)} + 1/2$$

and both types do not find it worthwhile to choose $s_i = 1$. ■

A.2 Proof of Proposition 2

Proof. For (i), it follows directly by noting that $\Pr(s_i = 1|\hat{\pi})$ *is increasing in* $\hat{\pi}$. To prove (ii), we first provide foundations for DD_j.

Foundations: Computing DD_j: Denote by θ the fraction of informed customers, v_G the value of ethical transactions to the customer, v_B the value of unethical transactions to the customer, where $v_G > v_B$. For simplicity, we assume that customers have the same willingness to pay for ethical transactions. The expected value of transacting (for customers) is: $v(R) = Rv_G + (1-R)v_B$, with a reduced form demand function: $q_1(R=1) = v(R=1) = \theta v_G$ for the informed customers versus $q_2(R = Pr(t_i = G|.,.)) = v(R = Pr(t_i = G|.,.)) = (1-\theta)v(R)$ for the uninformed customers. Thus, the aggregate market demand for ethical transactions is

$$q_G(R,.) = \underbrace{\theta v_G}_{q_1} + \underbrace{(1-\theta)v(R)}_{q_2}$$

Similarly, the aggregate demand is $q_B(R,.) = \theta v_B + (1-\theta)v(R)$ for unethical transactions.

Effects: Letting DD_j equal the aggregate demand q_{t_i}, and observing that $\frac{\partial q_{t_i=G}}{\partial \theta} = v_G - v(R) = v_G - Rv_G - (1-R)v_B \geq 0|_{R=1}$ in equilibrium. For unethical transactions, $\frac{\partial q_{t_i=B}}{\partial \theta} = v_B - v(R) \leq 0|_{R=1}$. We thus have the following result: For (ii), $DD_j(\theta)$ *is weakly-increasing in θ for G-types, but strictly decreasing in θ for B-types. Since $\hat{\pi}^*$ is decreasing in DD_j, noting that* $\lim_{DD_j \to +\infty} \hat{\pi}^* = 0$, *it follows that* $\Pr(s_i = 1)$ *is more likely for the G-types but less likely for B-types.* ■

B Balance and Attrition

Table 10: **Balance test I: representativeness of select-sample with market population (vendors)**

Supply side: Vendors	Constant	Select
Demographic Characteristics		
Female	0.398***	0.021
	(0.049)	(0.076)
Married	0.205***	0.083
	(0.043)	(0.065)
Akan ethnic	0.571***	8.96e-04
	(0.054)	(0.076)
Age	26.456***	0.716
	(0.585)	(1.117)
Education (any)	0.725***	-0.040
	(0.050)	(0.076)
Self employment	0.552***	-0.126*
	(0.058)	(0.075)
M-Money training	0.493***	0.043
	0.050	(0.070)
Poverty Indicators		
Household head read English	4.104***	0.102
	(0.163)	(0.223)
Outer wall used cement	3.909***	-0.306
	(0.222)	(0.342)
Toilet facility	4.617***	-0.349
	(0.140)	(0.268)
Number working mobile phones	8.466***	0.366
	(0.208)	(0.261)
Own working bicycle/ motor bicycle / car	1.554***	0.715
	(0.287)	(0.499)
Market: Size + Sales		
M-Money: Total volume [GHS] (daily)	2296.046***	24.611
	(129.932)	(178.263)
Non M-Money: Number customers (daily)	32.829***	-0.023
	(1.796)	(2.520)
Non M-Money: Total volume [GHS] (daily)	156.404***	-0.726
	(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 or market level. Each row is a separate regression. The F and Chi-squared tests are conducted excluding all market outcomes. Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

Table 11: **Balance test i: representativeness of select-sample with market population (customers)**

Demand side: Customers

	Constant	Select
Demographic Characteristics		
Female	0.628***	-2.0e-3
	(0.022)	(0.026)
Married	0.517***	0.021
	(0.019)	(0.024)
Akan ethnic	0.623***	-2.7e-3
	(0.036)	(0.039)
Age	38.635***	1.688*
	(0.737)	(0.891)
Education (any)	0.890***	9.7e-3
	(0.015)	(0.016)
Self employment	0.665***	0.025
	(0.029)	(0.029)
M-Money registered	0.905***	1.2e-3
	(0.014)	(0.017)
Poverty Indicators		
Household head read English	3.428***	-0.124
	(0.114)	(0.152)
Outer wall used cement	3.664***	-0.272
	(0.196)	(0.195)
Toilet facility	4.372***	-0.584
	(0.137)	(0.182)
Number working mobile phones	7.151***	-0.159
	(0.123)	(0.159)
Own working bicycle/ motor bicycle / car	1.180***	0.238
	(0.143)	(0.176)
Subjective Assessment: Fraud or Misconduct		
Attempted fraud experience (any)	0.611***	-0.041
	(0.040)	(0.039)
Ever over-charged/ unauthorized account use	0.292***	0.013
	(0.024)	(0.028)
Market: Features + Transactions		
Distance to closest formal bank (meters)	286.079***	147.891
	(73.105)	(107.315)
Distance to closest M-Money (meters)	66.295***	-10.758
	12.787	(13.021)
M-Money: Total use volume [GHS] (weekly)	129.227***	29.280
	(12.982)	(19.406)
Non M-Money: Number use (weekly)	2.062***	0.430
	(0.531)	(0.782)
Non M-Money: Total use volume [GHS] (weekly)	46.149*	-0.449
	(24.141)	(25.959)
Borrowing + Savings		
Likelihood to borrow via M-Money (1-5 scale)	1.515***	-0.065
	(0.073)	(0.069)
Likelihood to save via M-Money (1-5 scale)	2.126***	4.55e-3
	(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 vendor or market level. Each row is a separate regression. The F and Chi-squared tests are conducted excluding all market outcomes. Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

Table 12: **Balance test II: pre-intervention treatment-control differences (vendors)**

Supply side: Vendors

	Constant	**PT**	**MR**	**Joint: PT + MR**
Demographic Characteristics				
Female	0.551***	-0.180	-0.255*	-0.058
	(0.118)	(0.159)	(0.153)	(0.159)
Married	0.389***	-0.037	-0.202	-0.131
	(0.117)	(0.160)	(0.145)	(0.153)
Akan ethnic	0.491***	0.218	-0.118	0.189
	(0.119)	(0.156)	(0.161)	(0.151)
Age	27.097***	-0.413	2.163	-1.358
	(1.955)	(2.973)	2.845	(2.454)
Education (any)	0.697	-0.044	0.042	-0.041
	(0.126)	0.169	(0.165)	(0.163)
Self employment	0.443	0.058	0.008	-0.124
	(0.118)	(0.163)	(0.163)	(0.151)
M-Money training	0.340	0.265	0.293	0.170
	(0.119)	(0.163)	(0.159)	(0.160)
Poverty Indicators				
Household head read English	4.248***	-0.213	-0.093	0.139
	(0.295)	(0.506)	(0.480)	(0.4178)
Outer wall used cement	3.783***	0.038	-0.204	-0.486
	(0.591)	(0.790)	(0.794)	(0.784)
Toilet facility	4.464***	0.400	-0.581	-0.530
	(0.370)	(0.561)	(0.679)	(0.560)
Number working mobile phones	8.854***	-0.089	0.383	-0.346
	(0.276)	(0.490)	(0.490)	(0.449)
Own working bicycle/ motor bicycle / car	2.037***	0.004	0.359	0.483
	(0.642)	(1.072)	(1.002)	(1.052)
Poverty rate (Schneider 2015)	5.326	5.299	2.299	4.821
	(3.270)	(6.184)	(4.116)	(4.219)
Market: Size + Sales				
M-Money: Total volume [GHS] (daily)	1925.800***	305.049	478.480	665.939
	(555.950)	(789.582)	(902.508)	(1654.237)
Non M-Money: Number customers (daily)	32.473***	-2.080	-8.057	10.789
	(6.788)	(9.202)	(8.859)	(14.256)
Non M-Money: Total volume [GHS] (daily)	163.750***	-30.789	-14.096	14.986
	(61.630)	(66.831)	(69.562)	(73.869)
Joint F-test (linear), *p*-value			0.711	
Chi-squared test (probit), *p*-value			0.534	

Note: Observations are at the vendor or market level. Each row is a separate regression. The F and Chi-squared tests are conducted using the meta indicator **1(Information Assignment)** as the outcome and excluding all market outcomes. Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

Table 13: **Balance test II: pre-intervention treatment-control differences (customers)**

Demand side: Customers

	Constant	**PT**	**MR**	**Joint: PT + MR**
Demographic Characteristics				
Female	0.635***	0.003	-0.001	-0.034
	(0.053)	(0.061)	(0.069)	(0.064)
Married	0.505***	0.038	0.004	0.077
	(0.039)	(0.048)	(0.051)	(0.056)
Akan ethnic	.548***	0.101	0.077	0.092
	(0.072)	(0.092)	(0.102)	(0.090)
Age	39.380***	2.189	0.436	0.818
	(1.370)	(1.987)	(1.932)	(1.754)
Education (any)	0.891***	0.035	-0.027	0.021
	(0.025)	(0.029)	(0.042)	(0.033)
Self employment	0.668***	0.015	0.039	0.030
	(0.041)	(0.054)	(0.067)	(0.060)
M-Money registered	0.896***	-0.010	0.017	0.019
	(0.029)	(0.044)	(0.037)	(0.036)
Poverty Indicators				
Household head read English	3.353***	-0.081	-0.345	0.226
	(0.212)	(0.321)	(0.347)	(0.305)
Outer wall used cement	3.315***	-0.263	0.245	0.307
	(0.456)	(0.551)	(0.520)	(0.560)
Toilet facility	4.206***	-0.427	-0.478	-0.634*
	(0.169)	(0.377)	(0.405)	(0.327)
Number working mobile phones	7.086***	-0.415	-0.005	0.072
	(0.204)	(0.298)	(0.315)	(0.300)
Own working bicycle/ motor bicycle / car	1.141***	0.124	0.395	0.503
	(0.284)	(0.372)	(0.372)	(0.414)
Poverty rate (Schneider 2015)	11.280***	2.772	1.704	0.046
	(1.478)	(2.420)	(2.191)	(1.976)
Subjective Assessment: Fraud or Misconduct				
Attempted fraud experience (any)	0.565***	-0.000	0.018	2.41e-16
	(0.044)	(0.070)	(0.065)	(0.067)
Ever over-charged/ unauthorized account use	0.336***	-0.067	-0.037	-0.010
	(0.041)	(0.057)	(0.056)	(0.056)
Market: Features + Transactions				
Distance to closest formal bank (meters)	249.470**	-33.832	242.196	447.365*
	(96.807)	(127.385)	(255.640)	240.233
Distance to closest M-Money (meters)	45.623***	28.577	5.426	2.920
	(15.154)	(22.952)	(19.682)	(17.788)
M-Money: Total use volume [GHS] (weekly)	158.005***	-28.246	-9.495	37.712
	(35.465)	(40.296)	(41.623)	(55.060)
Non M-Money: Number use (weekly)	2.141***	-.255	1.049	0.532
	(0.606)	(0.748)	(1.972)	(1.230)
Non M-Money: Total use volume [GHS] (weekly)	26.706**	31.607	20.569	17.800
	(12.093)	(28.309)	(19.784)	(20.181)
Borrowing + Savings				
Likelihood to borrow via M-Money (1-5 scale)	1.391***	-0.011	0.098	0.130
	(0.120)	(0.141)	(0.171)	(0.174)
Likelihood to save via M-Money (1-5 scale)	2.103***	-0.070	0.087	0.085
	(0.177)	(0.248)	(0.246)	(0.264)
Joint F-test (linear), p-value			0.850	
Chi-squared test (probit), p-value			0.846	

Note: Observations are at the vendor or market level. Each row is a separate regression. The F and Chi-squared tests are conducted using the meta indicator **1(Information Assignment)** as the outcome and excluding all market outcomes. Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.10.

Table 14: **Attrition**

	PT	**MR**	**Joint: PT + MR**	**Control**	Total	Attrition
CENSUS (*Joint baseline*)						
Vendors					333	
Customers					1,921	
Markets (vendor×customers)					333	
SELECT SAMPLE (*Randomized*)						
Vendors	31	32	35	32	130	
Customers	272	257	276	185	990	
Markets (vendor×customers)	31	32	35	32	130	
ENDLINE (*Follow-up*)						
Vendors	26	28	28	25	107	23
	(84%)	(88%)	(80%)	(78%)	(82%)	(18%)
	(SD=37%)	(SD=33%)	(SD=40%)	(SD=42%)	(SD=38%)	(SD=38%)
Customers	230	207	230	143	810	180
	(85%)	(81%)	(83%)	(77%)	(82%)	(18%)
	(SD=36%)	(SD=39%)	SD=37%)	(SD=42%)	(SD=39%)	(SD=39%)
Markets (vendor×customers)	26	28	28	25	107	23
	(84%)	(88%)	(80%)	(78%)	(82%)	(18%)
	(SD=37%)	(SD=33%)	SD=40%)	(SD=42%)	(SD=38%)	(SD=38%)

Note: Table reports the summary statistics for the subsample that was successfully reached for a follow-up and for the subsample that was not successfully reached in endline phone surveys and manual contact tracing. Shown for both sides of the market (vendors versus customers). Attrition for endline audit exercises is 0.8%: 129 out of the 130 representative vendors were reached. There was only one unreachable vendor in the joint PT + MR program.

C Descriptive Statistics

Table 15: **Summary statistics of relevant variables from the market census**

	Vendors		Customers	
	Mean	SD	Mean	SD
Demographic Characteristics				
Female	0.398	0.489	0.623	0.484
Self employment	0.479	0.499	0.681	0.466
Self income -- monthly [GHS]	2.014	1.483	1.376	0.868
Married	0.249	0.432	0.535	0.498
Akan ethnic	0.572	0.494	0.621	0.485
Age (years)	26.291	8.242	39.545	15.021
Education (any)	0.691	0.461	0.896	0.304
M-Money training	0.508	0.500		
M-Money registered (self + any close person)			0.905	0.293
Poverty Indicators				
Household size (above 5)	0.223	0.416	0.244	0.430
Household head read English	0.769	0.421	0.606	0.488
Outer wall used cement	0.749	0.433	0.705	0.456
Toilet facility	0.891	0.311	0.849	0.357
Working mobile phone(s)	0.976	0.152	0.976	0.151
Own working bicycle/ motor bicycle/ car	0.280	0.449	0.214	0.410
Market: Access + Transactions + Sales				
Doing business experience (years)	2.051	2.12		
Joint venture: M-Money + other services	0.752	0.431		
M-Money: Total volume [GHS] (daily)	2260.569	3775.947		
Non M-Money: Number customers (daily)	32.791	47.067		
Non M-Money: Total volume [GHS](daily)	155.156	164.574		
Distance to closest formal bank (meters)			338.577	751.370
Distance to closest post office (meters)			382.932	250.737
Distance to closest M-Money (meters)			61.288	94.928
Formal bank user (of nearby banks)			0.806	0.395
Post-office user (of nearby offices)			0.092	0.290
M-Money user (of nearby vendors)			0.946	0.224
M-Money: Total use volume [GHS] (weekly)			144.199	396.283
Non M-Money: Number use (weekly)			2.272	14.766
Non M-Money: Total use volume [GHS] (weekly)			44.700	505.107
Borrowing + Savings				
Likelihood to borrow via M-Money (1-5 scale)			1.477	0.877
Likelihood to save via M-Money (1-5 scale)			2.112	1.213
Subject Assessment: Fraud or Misconduct				
Attempted fraud experience (any)			0.589	0.492
Ever over-charged			0.191	0.403
Ever over-charged + unauthorized account use			0.293	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 3: **Distribution (histogram) of total transactions at endline**

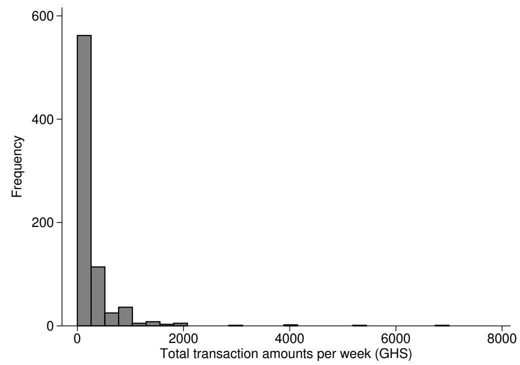


Figure 4: **Phone calls and reachability of subjects**

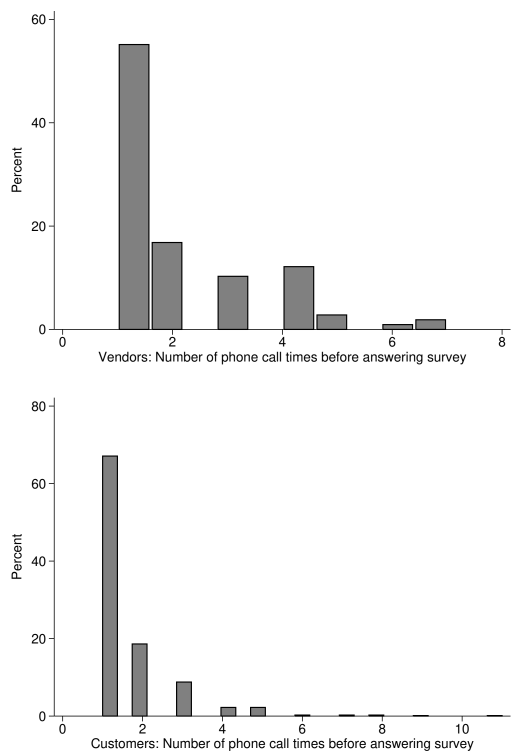
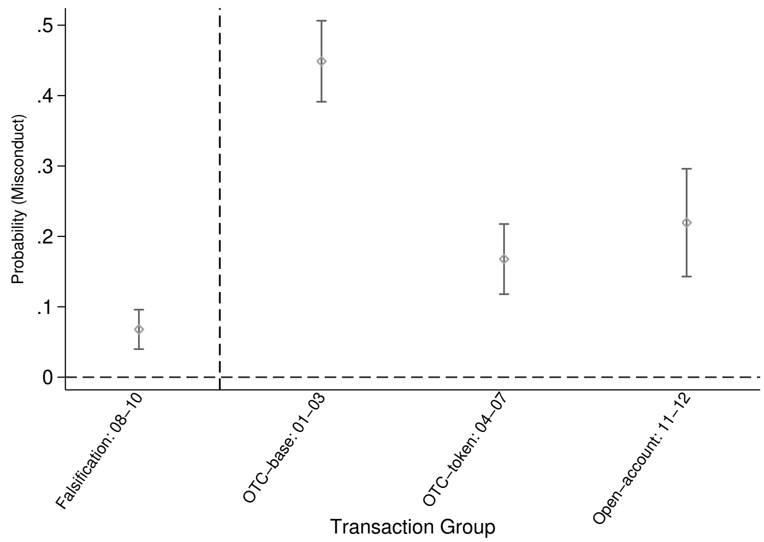


Table 16: **Misconduct at baseline: descriptive statistics based on transactional audit exercise, details**

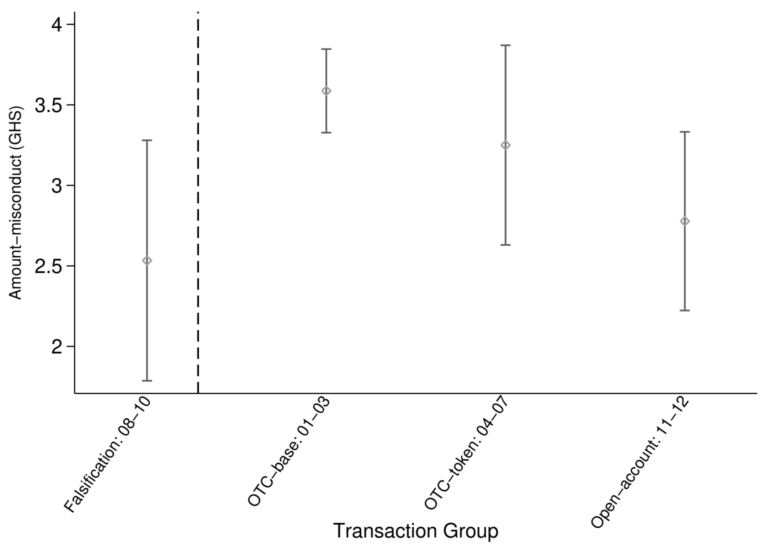
# Transaction type (description)	Outcome variable	Mean	SD	Transaction group	Mean	SD
01 Cash-in GHS50 - to others wallet	**1**[Misconduct=Yes]	0.35	0.480		0.44	0.498
	Overcharged [GHS]	4.65	1.093		3.58	1.498
02 Cash-in GHS160 - to others wallet	**1**[Misconduct=Yes]	0.52	0.502	$\{= \quad OTC - base$		
	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]	0.22	0.419		0.22	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 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 (i.e., more room for vendors to overcharge OTCs). **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 22% [SD=0.419] and the average overcharged-amount due to misconduct is GHS3.32 [SD=1.591], which represents $\frac{3.32}{4.03} \times 100 = 82\%$ of the average "official charge" for the transactional amounts used in the audit exercises. 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.

Figure 5: **Misconduct at baseline: descriptive statistics based on transactional audit exercise, graphical**



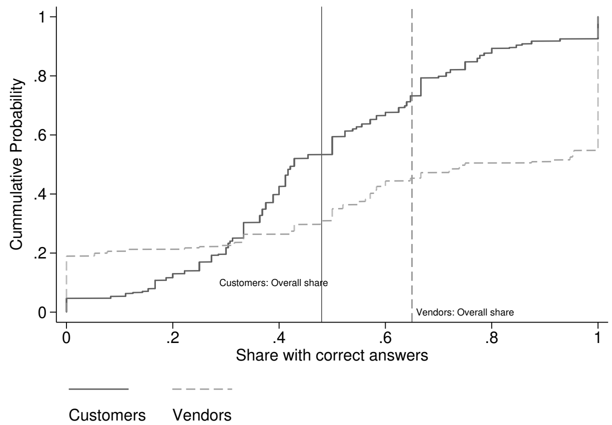
(a) **MISCONDUCT INCIDENCE × TRANSACTION GROUP**



(b) **MISCONDUCT SEVERITY × TRANSACTION GROUP**

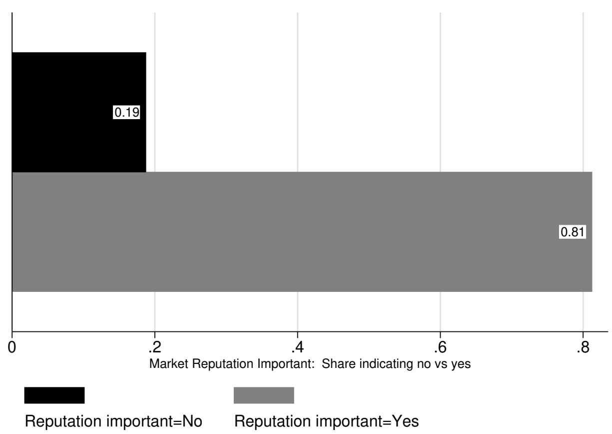
Note: Figures display the distribution of misconduct, measured as either the probability of the vendor committing a misconduct "incidence" (Figure (a)) or the amount overcharged as result of misconduct "severity" (Figure (b)) using actual transactional exercises at baseline. Transactions are categorized 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 in each transaction group are reported in the Table 16. 90% confidence intervals (CI) are displayed around the estimates. 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 6: **Asymmetric information about transactional prices**



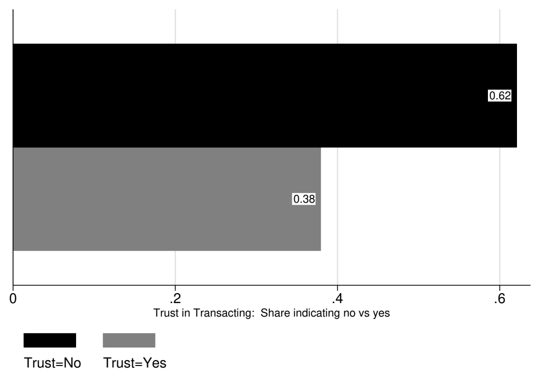
Note: Figure plots the distributions (CDFs) of the share of subjects with accurate answers for charges on randomly selected popular transactions (GHS200; GHS1200) derived with reference to their official or mandated rates (2GHS; 10GHS, respectively). A subject is correct if his/her answer matches the mandated rate. Shown separately for customers and vendors. From a Kolmogorov–Smirnov test for the equality of distributions, p-value < 0.01.

Figure 7: **Importance of reputation to vendors**



Note: Figure plots the share of vendors that value good market reputation through their money market transactions. Subjects were asked to indicate how important it is to show a high degree of good market image and responsibility to potential customers when carrying out M-Money transactions on a scale of 1 (not important) to 5 (very important). To ease the exposition, we first obtain the frequency distribution of the 1-5 value data and then find the median value (i.e., 4). All values above the median are recoded to be "yes" (reputation important), and those below are recorded as "no" (reputation not important). From an unpaired t-test for equality of vendors proportions of reputation-important and reputation-not important, p-value $= 0.000$.

Figure 8: **Consumer trust in performing money transactions at vendor points**



Note: Figure plots the share of customers, at baseline, that trust or not the money transactions they make at vendor banking points. Subjects were asked to indicate their level of trust for carrying out M-Money transactions at the vendor points from a scale of 1 (low) to 5 (high). To ease the exposition, we first obtain the frequency distribution of the 1-5 value data and then find the median value (i.e., 3). All values strictly above the median are recoded to be "yes" for trust in transacting (trust), and those below are recorded as "no" (distrust). From an unpaired t-test for equality of customers proportions of distrust and trust, p-value = 0.000.

Figure 9: **Misperceived beliefs about misconduct**

Note: Figure plots the share of transactions that are actually overcharged (truth) versus customers estimate of the share that are overcharged (perceived). From an unpaired t-test for equality of true misconduct $(1-\pi)$ and perceived misconduct $(1-\hat{\pi})$, p-value =0.000. π = the share of non-misconduct or transactions not-overcharged.

D Further Results: Treatment Effects

Table 17: **Effect of treatment on usage**

Model: Tobit
DV: Total Transaction per week (GHS)

	(1)	(2)	(3)	(4)
Treatment: Information Assignment (β)	116.628	106.077	99.402	95.292
	(52.439)	(52.149)	(53.718)	(52.489)
	[30.267, 202.989]	[20.194, 191.960]	[10.928, 187.875]	[8.840, 81.743]
sigma (σ)	581.695	576.667	571.064	563.983
	(83.946)	(83.240)	(83.464)	(82.838)
	[443.447, 719.942]	[439.580, 713.754]	[433.598, 708.529]	[427.547, 700.418]
Inference Robustness (β)				
Clustered S.E.	[18.033, 215.222]	[15.901, 196.253]	15.97649 182.828	[15.380, 175.203]
p-value: Permutation Test	0.069	0.085	0.085	0.091
Market District F.E.		X	X	X
Baseline usage			X	X
Controls				X
Observations	763	763	723	723
Mean of dependent variable (control)	198.956	198.956	198.956	198.956

Note: Market district is the randomization strata. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) and confidence sets (CS) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level.

Table 18: **Effect of treatment on usage**

Linear Model
DV: 0-1 Indicator for whether consumer is "using" M-Money

	(1)	(2)	(3)	(4)
Treatment: Information Assignment (β)	0.096	0.078	0.071	0.071
	(0.041)	(0.039)	(0.039)	(0.038)
	[0.028, 0.164]	[0.013, 0.143]	[0.006, 0.136]	[0.007, 0.133]
Inference Robustness (β)				
CI: Clustered S.E.	[-0.003, 0.197]	[0.010, 0.146]	[0.006, 0.136]	[0.011, 0.129]
CI: Wild Bootstrap	[0.028, 0.164]	[0.008, 0.135]	[0.002, 0.139]	[0.008, 0.132]
p-value: Permutation Test	0.017	0.028	0.045	0.049
p-value: L-S-X MHT Corr (2019)	0.022			
Market District F.E.		X	X	X
Baseline adoption			X	X
Controls				X
Observations	763	763	723	723
R-squared	0.008	0.074	0.075	0.105
Mean of dependent variable (control)	0.722	0.722	0.722	0.722
Lee (2009) Attrition Bounds				
Lower Bound:	0.083			
	(0.043)			
	[0.011, 0.154]			
Upper Bound:	0.142			
	(0.056)			
	[0.048, 0.234]			
Imbens and Manski (2004) CS	[0.025, 0.217]			
Behaghel et al. (2015) Attrition Bounds				
Lower Bound:	0.086			
	(0.041)			
	[0.005, 0.168]			
Upper Bound:	0.128			
	(0.041)			
	[0.047, 0.209]			

Note: Market district is the randomization strata. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) and confidence sets (CS) are reported in brackets. Reported confidence CIs for wild bootstrap and permutation tests derived from running 1000 replications in each case. Permutation test (for randomization inference) clustered at the (local) market level. L-S-X MHT Corr (2019) refers to the multiple hypothesis testing procedure presented in List, Shaikh and Xu (2019) for outcomes family (services usage; savings). 0-1 indicators for baseline migration motives (desire to migrate, plan to migrate, and permanent migration) used as predictors of attrition to tighten attrition bounds.

Table 19: **Gender differences in effects – usage, savings and poverty**

	Log Total Transaction per week (1)	Using M-Money (2)	Saving on M-Money (3)	PCA Index (1-3) (4)	Poverty Likelihood (5)
PANEL A	Pre-intervention gender gaps				
Female	-0.461	-0.003	-0.116	-0.060	0.399
	$(0.170)^c$	$(0.016)^c$	$(0.040)^c$	$(0.020)^c$	$(0.999)^c$
	$[-0.743, -0.179]^c$	$[-0.030, 0.023]^c$	$[-0.183, -0.048]^c$	$[-0.094, -0.026]^c$	$[-1.257, 2.056]^c$
Constant	32.603	0.931	0.499	-0.073	10.130
	$(0.452)^c$	$(0.049)^c$	$(0.105)^c$	$(0.050)^c$	$(1.436)^c$
	$[1.852, 3.353]^c$	$[0.848, 1.013]^c$	$[0.325, 0.573]^c$	$[-0.157, 0.010]^c$	$[7.748, 12.512]^c$
Observations	879	879	831	689	763
R-squared	0.170	0.064	0.148	0.133	0.113
PANEL B	Gender effects of treatment				
Treatment: Information Assignment (β)	0.118	0.027	0.029	0.105	3.282
	$(0.392)^c$	$(0.058)^c$	$(0.067)^c$	$(0.145)^c$	$(1.712)^c$
	$[-0.533, 0.769]^c$	$[-0.069, 0.124]^c$	$[-0.082, 0.142]^c$	$[-0.136, 0.347]^c$	$[0.442, 6.121]^c$
x Female	**0.454**	**0.068**	**0.073**	**0.161**	**-3.634**
	$(0.529)^c$	$(0.090)^c$	$(0.091)^c$	$(0.202)^c$	$(2.693)^c$
	$[-0.424, 1.332]^c$	$[-0.081, 0.218]^c$	$[-0.078, 0.226]^c$	$[-0.174, 0.497]^c$	$[-8.099, 0.831]^c$
Female	-0.774	-0.110	-0.144	-0.319	3.950
	$(0.479)^c$	$(0.081)^c$	$(0.081)^c$	$(0.179)^c$	$(2.202)^c$
	$[-1.569, 0.0214]^c$	$[-0.246, 0.025]^c$	$[-0.279, -0.009]^c$	$[-0.617, -0.022]^c$	$[0.298, 7.602]^c$
Observations	723	723	689	763	763
R-squared	0.109	0.106	0.069	0.113	0.123
Mean of dependent variable (control)	3.583	0.722	0.605	-0.198	10.186

Note: Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. c denotes clustering. Clustered standard errors (at local market level) reported in parenthesis. 90% confidence intervals (CI) are reported in brackets.

E Further Results: Belief Updates

Table 20: **Customers subjective belief about vendor responsibility (non-misconduct) increase at endline**

PANEL A			
Treatment: Information Assignment ($\gamma = \beta$)	0.065	0.069	0.067
	(0.035)	(0.035)	(0.035)
	[0.006, 0.122]	[0.011, 0.127]	[0.009, 0.126]
Baseline belief about $\mathbf{1}$(Non-misconduct=Yes)		X	X
Controls			X
Observations	943	941	941
Mean of dependent variable (control)	0.223	0.223	0.223
PANEL B			
Price Transparency ($\gamma \equiv \beta_1$)	0.064	0.067	0.066
	(0.042)	(0.042)	(0.042)
	[-0.005, 0.133]	[-0.001, 0.137]	[-0.004, 0.137]
Monitor and Report ($\gamma \equiv \beta_2$)	0.010	0.010	0.002
	(0.041)	(0.041)	(0.041)
	[-0.060, 0.076]	[-0.062, 0.073]	[-0.066, 0.069]
Joint program: PT + MR ($\gamma \equiv \delta$)	0.117	0.133	0.132
	(0.043)	(0.042)	(0.043)
	[0.047, 0.188]	[0.062, 0.203]	[0.061, 0.203]
Baseline belief about $\mathbf{1}$(Non-misconduct=Yes)		X	X
Controls			X
Observations	943	941	941
Mean of dependent variable (control)		0.223	
p-value (test: $\beta_1 = \delta$)	0.191	0.109	0.109
p-value (test: $\beta_2 = \delta$)	0.006	0.001	0.001
p-value (test: $\beta_1 = \beta_2$)	0.147	0.109	0.101
p-value (test: $\beta_1 + \beta_2 = \delta$)	0.435	0.301	0.272

Note: $\mathbf{1}(.)$ is a logical indicator that equals 1 if the argument in the parenthesis is true, 0 otherwise. In each market, we compute the baseline outcome as the share of experimental customers that indicate no experiences of misconduct (i.e., non-misconduct). Includes baseline outcomes, and additional controls. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) are reported in brackets. Panel A reports the meta estimate of treatment effects, while panel B shows the effects separately for each information program.

Table 21: **CUSTOMERS BELIEF UPDATE: EFFECT OF INFORMATION SETS ON CUSTOMERS CORRECT GUESS OF VENDOR IRRESPONSIBILITY (OR MISCONDUCT)**

$\mathbf{1}(\text{GUESS}_i\text{=AUDIT TRIAL}_j)$			
PANEL A			
Treatment: Information Assignment ($\gamma \equiv \beta$)	0.084	0.085	0.088
	(0.030)	(0.030)	(0.029)
	[0.033, 0.134]	[0.034, 0.136]	[0.039, 0.137]
Baseline belief about $\mathbf{1}$(Misconduct=Yes) $\approx \Pr(t_i)_{base,j}$		X	X
Controls			X
Observations	763	763	763
Mean of dependent variable (control)	0.101	0.101	0.101
PANEL B			
Price Transparency ($\gamma \equiv \beta_1$)	0.067	0.068	0.072
	(0.037)	(0.037)	(0.036)
	[0.007, 0.128]	[0.007, 0.129]	[0.012, 0.131]
Monitor and Report ($\gamma \equiv \beta_2$)	0.012	0.013	0.013
	(0.036)	(0.036)	(0.035)
	[-0.047, 0.071]	[-0.046, 0.072]	[-0.044, 0.071]
Joint program: PT + MR ($\gamma \equiv \delta$)	0.164	0.167	0.173
	(0.039)	(0.040)	(0.039)
	[0.098, 0.229]	[0.101, 0.233]	[0.109, 0.238]
Baseline belief about $\mathbf{1}$(Misconduct=Yes) $\approx \Pr(t_i)_{base,j}$		X	X
Controls			X
Observations	763	763	763
Mean of dependent variable (control)	0.101	0.101	0.101
p-value (test: $\beta_1 = \delta$)	0.011	0.010	0.009
p-value (test: $\beta_2 = \delta$)	0.000	0.000	0.000
p-value (test: $\beta_1 = \beta_2$)	0.112	0.114	0.093
p-value (test: $\beta_1 + \beta_2 = \delta$)	0.111	0.105	0.092

Note: $\mathbf{1}(.)$ is a logical indicator that equals 1 if the argument in the parenthesis is true, 0 otherwise. Includes randomization strata (market district) dummies, baseline outcomes, and additional controls. Robust standard errors reported in parenthesis. 90% confidence intervals (CI) are reported in brackets. Panel A reports the meta estimate of treatment effects, while panel B shows the effects separately for each information program.

F Anti-Misconduct Information Programs – Exhibits

F.1 FIRST: VISIT NEARBY CUSTOMERS

PREAMBLE: Greetings Madam/ Sir... My name is....
Please recall we visited your unit in February 2019 to do a survey of (the M-Money business) to find out (how customers, like you, understand the business of M-Money and other services

their centers provide). Today, we have come to provide additional education about M-Money for research and to help make the market better and understandable. You may call the research team anytime *if* in any doubt (Phone: XXXXXXXXXX) (omitted to preserve privacy).

F.1.1 T1 - PRICE TRANSPARENCY, PT

Our message is simple. We want to remind you that you should:

- Make sure to ask for official tariff sheets when transacting: e.g., opening new Wallet, OTC, sending. Simply ask.
- When opening a new Wallet don't pay fees – deposit should be credited to your account, check it right away.
- Example of common charges: (i) Pay 0.5GHC if putting 50GHC on someone's account; (ii) 1.6GHC if putting 160GHC on someone's account; 10GHC if putting 1100GHC on someone's account; (iv) it's free to put any money on your own Wallet.
- Research Officer, ask customers: (1) Repeat information provided. (2) Please how would you rate the usefulness of this information we provided for your financial decisions on a 5-point scale? [1=Not useful, 2=Quite useful, 3=Useful, 4=Very useful, 5=Extremely useful]. (3) Leave a hard (paper) copy of this information with subject.

F.1.2 T2 - (MARKET) MONITOR AND REPORT, MR

Our message is simple. We want to remind you that *if* you:

- Suspect any discrepancy or glitches as you make any M-Money transactions, you should call MTN fraud department on NUMBER (Toll-Free number: 100) to report it, right away.
- There is an MTN fraud department; **free** to call. They always help.
- Research Officer, ask customers: (1) Repeat information provided. (2) Please how would you rate the usefulness of this information we provided for your financial decisions on a 5-point scale? [1=Not useful, 2=Quite useful, 3=Useful, 4=Very useful, 5=Extremely useful]. (3) Leave a hard (paper) copy of this information with subject.

F.1.3 T3 - PT+MR

We have two main messages:

- First, we want to remind you that you should: Make sure to ask for official tariff sheets when transacting: e.g., opening new Wallet, OTC, sending. When opening a new Wallet don't pay fees – deposit should be credited to your account, check it right away. Example of common charges: (i) Pay 0.5GHC if putting 50GHC on someone's account; (ii) 1.6GHC if putting 160GHC on someone's account; 10GHC if putting 1100GHC on someone's account; (iv) it's free to put any money on your own Wallet.
- Second, we want to remind you that *if* you: Suspect any discrepancy or glitches as you make any M-Money transactions, you should call MTN fraud department on NUMBER (Toll-Free number: 100) to report it, right away. There is an MTN fraud department; **free** to call. They always help.
- Research Officer, ask customers: (1) Repeat information provided. (2) Please how would you rate the usefulness of this information we provided for your financial decisions on a 5-point scale? [1=Not useful, 2=Quite useful, 3=Useful, 4=Very useful, 5=Extremely useful]. (3) Leave a hard (paper) copy of this information with subject.

F.2 SECOND: VISIT REPRESENTATIVE OR SELECT VENDOR

PREAMBLE: Greetings Madam/ Sir... My name is....
Please recall we visited your unit in February 2019 to do a survey of (the M-Money business) to find out (how merchants, like you, understand the business of M-Money and other services that your centers provide). Today, we have come to provide additional education about M-Money for research and to help make the market better and understandable. You may call the research team anytime *if* in any doubt (Phone: XXXXXXXXXX) (omitted to preserve privacy).
[RESEARCH OFFICER: LET'S BLUFF ABOUT INTERVENTIONS GIVEN TO CUSTOMERS]: We have educated "nearby" customers in this locality about M-Money (since many of them don't understand M-Money's workings well) that:

F.2.1 T1 - PRICE TRANSPARENCY, PT

They should:

- Make sure to ask for official tariff sheets when transacting: e.g., opening new Wallet, OTC, sending.
- When opening a new Wallet don't pay fees – deposit should be credited to their account, check it right away

- Example of common charges: (i) Pay 0.5GHC if putting 50GHC on someone's account; (ii) 1.6GHC if putting 160GHC on someone's account; 10GHC if putting 1100GHC on someone's account; (iv) it's free to put any money on their own Wallet.
- Research Officer, ask vendors: (1) Please how would you rate the usefulness of this information we provided for your financial business on a 5-point scale? [1=Not useful, 2=Quite useful, 3=Useful, 4=Very useful, 5=Extremely useful]. (2) Leave a hard (paper) copy of this information with subject.

F.2.2 T2 - (MARKET) MONITOR AND REPORT, MR

- *If* they: Suspect any discrepancy or glitches as they make any M-Money transactions, they should call MTN fraud department on NUMBER (Toll-Free number: 100) to report it, right away.
- There is an MTN fraud department; **free** to call. They always help.
- Research Officer, ask vendors: (1) Please how would you rate the usefulness of this information we provided for your financial business on a 5-point scale? [1=Not useful, 2=Quite useful, 3=Useful, 4=Very useful, 5=Extremely useful]. (2) Leave a hard (paper) copy of this information with subject.

F.2.3 T3 - PT+MR

Two main messages:

- First, they should: Make sure to ask for official tariff sheets when transacting: e.g., opening new Wallet, OTC, sending. When opening a new Wallet don't pay fees – deposit should be credited to their account, check it right away. Example of common charges: (i) Pay 0.5GHC if putting 50GHC on someone's account; (ii) 1.6GHC if putting 160GHC on someone's account; 10GHC if putting 1100GHC on someone's account; (iv) it's free to put any money on their own Wallet.
- Second, *if* they: Suspect any discrepancy or glitches as they make any M-Money transactions, they should call MTN fraud department on NUMBER (Toll-Free number: 100) to report it, right away. There is an MTN fraud department; **free** to call. They always help.
- Research Officer, ask vendors: (1) Please how would you rate the usefulness of this information we provided for your financial business on a 5-point scale? [1=Not useful, 2=Quite useful, 3=Useful, 4=Very useful, 5=Extremely useful]. (2) Leave a hard (paper) copy of this information with subject.

G Vendor Banking Points – Photos

Figure 10: **Vendor banking points**

Note–Providers: MTN Mobile Money, AirtelTigo Money, Voda Cash