Misconduct and Reputation under Imperfect Information

Preprint · April 2023

1 author:

Francis Annan
University of California, Berkeley

Misconduct and Reputation under Imperfect Information

Francis Annan♣

University of California, Berkeley

MAY 2023

Abstract

Misconduct – market actions that are unethical and indicative of fraud – is a significant yet poorly understood issue that underlies many economic transactions. We design a field experiment to study the impact of two-sided anti-misconduct information programs, which we deploy on the local markets for mobile money (Human ATMs) in Ghana. The programs lead to large reduction in misconduct (-21pp=-72%) and as a result, broader improvements in overall market activity, consumer welfare, and firm revenue. We show the treatment effect is due to a combination of more accurate consumer beliefs about misconduct and increased vendor reputation concerns.

Keywords: *forensics and information* (**D83**), *vertical markets and reputation* (**L14**, Z13), *household finance* (D14, **O12**), *consumer protection* (**D18**), *firm behavior and growth* (**L26**, L13)

♣ UC Berkeley. e-mail: `fannan@berkeley.edu`. For helpful discussions and suggestions, we thank Iwan Barankay, Arun Chandrasekhar, Pascaline Dupas, Chiara Farronato, Mariassunta Giannetti (discussant), Xavier Gine, Paul Goldsmith-Pinkham, Marco Gonzalez-Navarro, Brett Green, Glenn Harrison, Sean Higgins, Dean Karlan, Supreet Kaur, Erzo F.P. Luttmer, Isaac Mbiti, Jeffrey Perloff, Matt Shum, Sandip Sukhtankar, Tavneet Suri (discussant), Catherine Tucker, Christopher Snyder, Christopher Udry, and several seminar participants at Household Finance Seminar, University of Virginia Economics, University of Pennsylvania-Wharton, Georgia State University, Michigan State University, Columbia University, UC Berkeley, Cornell University, Dartmouth College, PEDL 2020 Conference on Firms in Low-Income Countries, CSAE Oxford 2021, BREAD 2021 Conference on Behavioral Economics and Development, NBER 2021 Economics of Digitization Meeting, BREAD 2021 Conference on the Economics of Africa, and NBER 2021 Summer Institute. We thank the NBER Development Economics Program Fall 2020 meeting participants for their comments on an earlier draft. We are grateful to three anonymous referees and the editor for their insightful comments. Field support from Samuel Adotevi and Kwamena Arkafra (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

Casual empiricism suggests that firm misconduct—failure to comply with rules/ laws/standards—is prevalent and costly. As FORTUNE Magazine claims, "[Businesses] lie, they cheat, they steal, and they've been [mostly] getting away with it too long." (FORTUNE 2002). This assertion is at the heart of several regulatory apparatuses designed to prevent and punish firm misconduct, with leading, recent examples in markets for digital financial services (DFS). In 2018, the Global System for Mobile Communications (GSMA) launched a global certification program meant to bring safer, more transparent, and resilient financial services to millions of mobile money users around the world (GSMA 2018). Similar initiatives have emerged at national levels to increase government oversight of financial services provision and governance, including Ghana's DFS Policy of 2020 and Kenya's Payment Systems Act of 2011. These initiatives were particularly targeted at DFS for the poor, as firms in the marketplace recently came under serious scrutiny for issues pertaining to their conduct and to consumer protection (Garz et al. 2021).

Despite its evident importance, there are major gaps in research on firm conduct. In particular, the key ingredients for advancing basic and applied knowledge—mechanisms, cost of misconduct, and, especially, negative externalities from misconduct—are often unidentified (Zitzewitz 2012; Zinman and Zitzewitz 2016). This paper addresses these gaps in two ways. First, we collect original data on misconduct prevalence and severity, market participant beliefs, and firm and consumer outcomes. Second, we run a market-level field experiment, testing scalable approaches to enriching information sets and lowering enforcement costs.

We conduct our large-scale experiment in the retail market for mobile money (M-Money). This important financial market innovation has the potential to improve welfare and reduce poverty (Suri and Jack 2016; BMGF 2021). Tariffs on M-Money transactions are set *ex-ante* by the providers that contract with market vendors or sellers, who are not allowed to alter these rates. Nonetheless, these markets exhibit substantial vendor misconduct (vendors overcharge on over 22% of transactions), imperfect information about official tariffs (poor consumer knowledge relative to vendors), consumer mistrust (62% of customers distrust transacting at vendor points), and misperceived beliefs (upwardly-biased consumer beliefs) about misconduct. These features, which we show at baseline, make M-Money a relevant setting to study misconduct and reputation under imperfect information. Indeed, this form of seller misconduct or overcharging in payment markets exists in many other countries like Kenya, Uganda, and Nigeria (Blackmon, Mazer, and Warren 2021).

We construct a unique census of 130 independent, spatially-distinct local markets (localities or villages) across 9 districts for M-Money between February–March 2019, as detailed data about vendor and customer interactions are unavailable. The large number of distinct markets allows for randomization at the market level, as well as identification of both differentiated markets (with minimal cross-market spillovers) and spillovers therein. Markets designate pairs of vendors and their nearby customers. We perform our experiment by randomly assigning these markets to one of three anti-misconduct information programs: one about price transparency (PT), the second about monitoring and reporting (MR), and the third about both (PT+MR, their interaction). In the PT treatment, consumers receive relevant information and training about official transaction charges. In the MR treatment, consumers receive a toll-free number to report suspected misconduct to providers or authorities. The joint treatment combines the PT and MR treatments. In all cases, vendors are informed that customers have received such information, and the same information sets are then given to the vendors, making our interventions two-sided. Thus, the interventions empower consumers with technologies to enforce market vendors' trustworthiness by relying on social sanctions and/or punishment. For each locality, we apply the intervention to one random vendor and nearby customers. We track additional non-treated vendors to examine spillover effects.

We implement an audit study to measure vendor misconduct: trained auditors visit vendor points to make actual transactions. The transaction charges are then compared to the official tariffs to measure misconduct (Egan, Matvos, and Seru 2019; Annan 2020). Misconduct in markets remains a poorly understood issue due to the empirical difficulties in measuring it objectively. Here, we develop a procedure to cleanly measure misconduct connected to increased transaction costs and shrouded prices. Our dataset is unique due to its size (130 random vendors and 990 customers), the expansive set of outcomes from both sides of the market, the administrative audit measures of misconduct, market census and surveys, and the 2 x 2 random information variation at market level. We have five set of results.

First, the intervention reduces vendor misconduct and improves consumers' beliefs about vendor's honesty dramatically. Overall, the incidence of vendor misconduct decreases by -21 percentage points (pp) = -72%, while the severity of misconduct decreases by -GHS0.60 (-$0.14) = -78%. With a control mean of GHS0.70, the latter means the intervention leads the total fee (official charge + misconduct) to fall from about 1.70% to about 1.10%, implying a 40% reduction of typical M-Money transaction fees. Consumers' perception of honest vendor behavior increases (+7.0 pp = +30% overall), and, importantly, such beliefs are positively correlated (+27 pp = +51%) with the objective audit measure of misconduct, implying more accurate and updated consumer beliefs due to the information sets. The combined information intervention has the greatest reduction in market vendors' misconduct. However, the PT-alone and MR-alone programs also have meaningful impacts on misconduct.

Second, customers meaningfully increase their use of M-Money (+10% to +45%) and their likelihood of saving on M-Money (+7.5 pp =+12.1%) at vendor points. Third, vendors' sales revenue increases. Overall, the information programs significantly increase vendors' total sales (+52%). This result is consistent with the estimated consumer impact. Thus, reducing vendor misconduct can enhance the efficiency of local markets by increasing market activity. For context, a 45% increase in consumer demand (or 52% increase in vendors' sales) in response to a 40% total fee (official charge + misconduct) reduction is reasonable; it is an elasticity of about 1.1 (or 1.3) and falls within the range of market effects from relevant M-Money tax and subsidy policy experiments.

Fourth, we find significant spillover effects. Non-treated vendors located in treated localities reduce their misconduct by -21 pp overall, suggesting our information programs have market-wide behavioral impact. We estimate a 55% increase of vendors' non-M-Money business services revenue. Fifth, consumers in treated markets are -6.8 pp (7.6%) less likely to experience shocks they could not financially remedy. The combined program shows larger impacts across the various outcomes than the alternative individual information programs, suggesting that the two individual information sets complement one another. We do not find evidence of an impact on overall poverty levels, the number of customers, or business exits.

Why does everyone benefit from our market-level interventions? One possible explanation is that vendors face a prisoner's dilemma problem. Vendors (including other market participants) would be collectively better off if vendors did not cheat, but there are private benefits to deviating from a low-misconduct equilibrium, resulting in a privately profitable high-misconduct equilibrium. In this market with significant information frictions, it might be difficult to establish a reputation for low rates, which result in a better collective outcome. Thus, transparency and monitoring systems that enforce a low-misconduct equilibrium could be welfare improving. We develop a simple framework to evaluate reputation, where vendors expect that they are more likely to be perceived by customers as irresponsible if they commit misconduct in our experiment (Macchiavello and Morjaria 2015).

We make three main contributions to the existing literature. First, we contribute to the literature on information and business growth in developing countries. Previous studies have emphasized several barriers to business growth, including managerial constraints (Bloom et al. 2013), limited network and interfirm relations (Cai and Szeidl 2017), lack of capital (De Mel, McKenzie, and Woodruff 2008), lack of market access (Atkin, Khandelwal, and Osman 2017), and asymmetric information (Jensen and Miller 2018; Bai 2019). Here, we emphasize miscalibrated consumer beliefs about seller misconduct and vertical market structure as potential barriers. Our market operates on a vertical structure: service providers at the upstream set up commission-motivated vendors at the downstream who are not allowed to alter rates, leading to a "version" of the well-known single versus double marginalization problem (Tirole 1988, Chapter 4; Janssen and Shelegia 2015). The treatment pushed the double marginalization high price (high misconduct) to a single marginalization lower price (low to no misconduct). Thus, we show that all players on the vertical structure—providers, vendors, and customers—can be made better off under the single marginalization result, which is novel and interesting.

Second, we add to the literature on forensic economics (see e.g., Olken and Pande 2012; Zitzewitz 2012 for detailed reviews). Misconduct underlies many economic and financial transactions (Egan, Matvos, and Seru [2019, 2022]; Annan 2020), yet the sources of such concealed behavior are not well-understood. We emphasize how imperfect information might exacerbate misconduct, showing in our experiment that providing symmetric information to transacting parties raises vendor concerns for reputation. Little is known about how reputational losses discipline business misconduct (Karpoff 2012 provides a review indicating ambiguous effects). We emphasize how local sanctions via reputation-building can promote rural financial institutions and development in low-income settings (see Munshi 2014 for a review).

Third, we contribute to the literature on information disclosure, household finance, and financial technology adoption. There is much existing research on the consumer effects of FinTech (Jack and Suri 2014; Suri and Jack 2016), but there is almost no work on supply-side behavior (Higgins 2020). We emphasize seller misconduct as a key barrier to both sides of the market, and show that reducing it via information disclosure has broader impacts on consumers and businesses. We show that disclosure – transparency and monitoring – is beneficial to retail businesses and improves sales revenue (Brown, Hossain and Morgan 2010). Moreover, we document misconduct in payment markets, which is an open—and high-priority— area of research, particularly in developing countries,[1] where consumers lack

[1] Hasanain et al. (2023) discloses information about artificial insemination services of livestock provision to farmers in Punjab, Pakistan, through an information clearinghouse. Unlike Hasanain et al. (2023), we (i) deploy market-level interventions and set up a design that allows us to measure within-market spillovers; (i) have direct measures of consumers' subjective beliefs and objective measures of vendor misconduct, which allows us to define and evaluate belief updates, bias vs price effects, and

experience with FinTech (Garz et al. 2021), and higher transaction fees can act as a barrier to the adoption of payment services (Higgins 2020), as well as reduce risk sharing across households (Jack and Suri 2014). Our study is the first, to our knowledge, to provide quantitative estimates of both seller misconduct in digital financial markets and the value of anti-misconduct information programs, particularly in environments where M-Money has the potential to reduce poverty and meaningfully improve the welfare of consumers.

From a policy perspective, our results highlight how the provision of low-cost, two-sided information might influence vendor conduct and consumer trust, 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 buyers are, and providing information about price transparency and monitoring to both sides of the market, could potentially be used to build trust and increase the benefits of emerging markets for digital finance.

We proceed as follows: In Section II, we describe the research setting, and in particular, vendor misconduct within M-Money. Section III contains the description of our experimental design and data. Section IV presents our main results. In Section V, we discuss the implications of our results, and describe the framework we use to derive our preferred interpretation of the results. We conclude the paper with Section VI.

II Research Setting

A. Mobile Money

M-Money provides financial services that are delivered on digital mobile networks to consumers. The market for M-Money comprises (i) service providers, (ii) vendors, and (iii) customers. In Ghana, there are four providers: MTN M-Money, Vodafone VodaCash, AirtelTigo Money, and GCB Ltd.'s G-Money. MTN has about 90% share of the market. Providers are joint partnerships between mobile network operators (MNOs) and commercial banks. Mar-

reputation; and (iii) are able to measure broader impacts on both prices, quantities, and welfare.

ket vendors (or sellers) are small business retail distribution points and correspond to outlets, shops, premises, or local banking channels. They conduct M-Money transactions on behalf of the providers.

Vendors register new accounts (also called "wallets") for customers and act as cash-in (deposits, transfers) and cash-out (withdrawals) transaction points for customers (i.e., Human ATMs).[2] Vendors can freely enter and exit the market. To establish the retail business of M-Money, vendors must have the required documentation and meet certain structural and monetary requirements. Vendors should have a permanent space from which to operate and a minimum startup capital of GHS4000 ($US781.25)[3], which we observe in practice can be relaxed, depending on the environment. All vendors must undergo official business training about the tariffs, commissions, and other services. They generally earn transactional commissions on sales revenue as their profit. In comparison, customers receive little to no information about M-Money's transaction tariffs and services when they sign up. The tariffs on transactions at vendor points are set *ex-ante* by providers, so market vendors are not allowed to marginalize. Thus, the M-Money setup has a vertical market structure: service providers at the upstream set up vendors at the downstream, who work for them and earn commissions on sales.

The introduction, and significant penetration, of digital mobile telecommunications has provided a cheap infrastructure to make M-Money services accessible even to poor and low-income societies. In these environments, formal financial institutions are shallow and largely absent (see Banerjee and Duflo [2006; 2011] for authoritative surveys), making M-Money a competitive financial option. Evidence suggests that M-Money has the potential to reduce poverty and improve the welfare of consumers in Sub-Saharan Africa and Asia through several channels (Jack and Suri 2014; Suri and Jack 2016; BMGF 2021). M-Money

[2]There is demand for vendors because users (i) are poorly informed about how to operate M-Money platform's menu for self-serve transactions, or (ii) have to make deposits and open new accounts, or (iii) want to avoid digital taxes and digital loan defaults, which only apply to self-serve transactions in our setting, or/ and (iv) otherwise direct merchant payments are limited as merchants mostly accept only cash for goods and services. As in our setting, about one-third of consumers in Tanzania, Uganda, and Bangladesh cannot do their own M-Money transactions and tend to rely on vendors (TCI: Transaction Cost Index 2023).

[3]MTN Mobile Money 2021: https://mtn.com.gh/momo/agent/

is an important market, but could be constrained by market misconduct that shrouds prices and increases transaction costs. Providers at the upstream have limited oversight into the behavior of downstream vendors, and consumers in low-income environments are poorly informed.

Similar to other banking and financial services, the business of M-Money likely faces fraud and misconduct, which could take different forms. Indeed, vendor misconduct is widespread. Recent surveys from Innovations for Poverty Action (IPA) compare market misconduct (overcharging of services) in Uganda, Nigeria, and Kenya (Blackmon, Mazer, and Warren 2021): 33%, 42%, and 3% of consumers respectively reported vendor overcharging. For our experimental sample, this will correspond to 22% of transactions being overcharged or subject to unofficial fees. In policy circles, regulators from Bank of Ghana, for example, have expressed concerns about such potential market misconduct. MNOs and their commercial partners have been asked to build more risk- and fraud-resilient financial infrastructures.[4] Our present study is designed to understand misconduct at retail vendor points (see Figure D.1 in Appendix D), as well as the effect of social sanctions and/or punishment, and to evaluate their potential market-wide impacts. We do this in a rural context where the business of M-Money could have larger positive impacts, if well designed.

B. Descriptive Motivating Facts

We document several facts about our setting, and, in particular information, frictions and vendor misconduct, drawn from a pre-experiment market census in Eastern Ghana. Detailed vendor $\times$ customer data on M-Money is unavailable, so, between February and March 2019, we carry out a census of the market for M-Money, spanning nine districts. Districts are made up of sub-administrative units called "localities" or villages. The select localities have a mean and median population of 3900 and 2300 people respectively as of 2018. We use a master gazetteer of localities kept by the Ghana Statistical Service. Our census exercise

[4] "*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.

successfully documents the universe of all vendor points and surrounding households (within a five-house radius around a given vendor) across 130 localities (Figure B.1 displays the spatial distribution). This yields a total of 333 vendors and 1,921 customers or households. We focus on nearby households in order to maximize our chances of studying households that might make transactions with select vendors, while also minimizing costs. We define a local market as the pair: vendor $\times$ the set of all nearby customers.

We gather information about basic demographics, poverty and assets, and detailed market records on M-Money and non-M-Money services, including general to specific knowledge of vendors and consumers about M-Money transactions. We also obtain additional household information from customers on personal finance, shocks, and investments. Detailed summaries are available in Annan (2020) and upon request. Table B.8 shows summary statistics for the market. Female vendorship is 39%, meaning that these local markets are disproportionately made up of male vendors. Of potential customers, 62% are females, and customers are more likely than vendors to be self-employed, married, and older. The overwhelming majority (90% [SD=0.29]) of customers, as well as their networks of close family and friends, have registered for a M-Money account, indicating that it is likely a popular financial technology.

We turn next to specific features of the market. With an average experience of two years doing M-Money business, a vast majority (75% [SD=0.43]) of vendors operate as a bundled store, bundling M-Money with other services.[5] The average daily sales per vendor for M-Money is about GHS2,260 (US$442). With an official sales commission of 1%, the average vendor will earn a daily profit of around GHS23. The majority of households use M-Money services rather than other alternative commercial financial services: 95% of customers are M-Money users, 80% are past formal bank users, while just 9% are post office users. This can be explained by the convenient access and arguably lower charges of M-Money, and by the relative inaccessibility and distance of other services. We use the census to document

[5]We identified bundled services including 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, electronics and accessories, photocopying and typesetting, educational/online results checking, and prepaid electric credit, among others. Baseline sales revenue from these non-M-Money services represents about 7% of the sales revenue from M-Money (Table B.8).

three facts that suggest information frictions matter.

Fact 1: There is high vendor misconduct, but customers misperceive misconduct.

Figure B.10 compares true versus subjective beliefs of misconduct. Our audit transactions provide an objective (true) misconduct incidence of 22% [SD=0.41, n=663] at vendor points, which is high. We ask customers, at baseline, whether they believe they have experienced overcharges at vendor points (the incidence of misconduct), yielding an overall subjective incidence of 59% [SD=0.49, n=1921]. This suggests that consumers misperceive vendor misconduct (upwardly-biased consumer beliefs).

Fact 2: High asymmetric information about official prices between vendors and customers.

In a series of tests, both vendors and customers are asked to indicate the official charges for two randomly chosen transactions of sizes GHS200 (small to medium) and GHS1200 (large). This provides us an estimate of their knowledge about the official charges. We are careful to inform vendors at the beginning that we are not there to perform any actual transactions, but rather to assess their overall knowledge of the market. Knowledge tests are taken towards the end of the surveys for both sets of subjects. Results are displayed in Figure B.7, showing strong evidence of asymmetric information: vendors have superior knowledge of official prices relative to customers. This creates opportunities for vendor misconduct. Overall, consumers are correct 42% (median) of the time, while vendors are correct 80% (median) of the time (an incentivized measure increases vendor accuracy to 95%, without any change to consumer accuracy). These results are expected, because, unlike customers, vendors receive formal training before they start their businesses.

Fact 3: Customers mistrust vendors, but vendors value good reputation.

We solicit information about customers' level of trust in vendors when carrying out their transactions. Figure B.9 reports the results, suggesting limited level of trust. About 62%

[SD=0.48, n=1275] of customers indicate distrust in transacting at vendor points, while the rest (38% [SD=0.48, n=779]) indicate trust. We ask a random sample of vendors about the importance of demonstrating good market reputation (or image and responsibility) to potential customers through their market transactions. As shown in Figure B.8, the vast majority of vendors (81% [SD=0.391]) consider good market reputation or image as important, suggesting there is likely a positive return to vendors for good market reputation, if they are viewed by customers as responsible, though this may be constrained by the limited consumer trust. Vendors have poor reputation in the market, perhaps because customers are unable to infer vendors' behavior.

Together, our markets reflect a setting where (i) misconduct is high; (ii) consumers are uninformed; (iii) vendors value their reputation in the market, but good reputation is difficult to establish, because consumers, not knowing official prices, cannot determine whether vendors are being honest; and (iv) consumers underperceive the level of vendors' honesty (upwardly-biased beliefs about vendor misconduct). The results demonstrate that information matters and there is room to build trust and reputational capital in the market.

But *why the high misconduct at baseline?* We advance 9 separate hypotheses to shed light on why the pre-experiment market equilibrium may have so much misconduct. We implement follow-up surveys to gather various views from both managers of the service provider and vendors in control markets. Managers and vendors were invited to rank 9 hypotheses in order of the most plausible reason for "why vendor misconduct is prevalent in low-income areas". Figure B.11 shows the results, separately, for managers (n=29) and vendors (n=58). The top 4 ranked hypotheses are (i) poorly-informed consumers about prices and redress channels, (ii) too low vendor commissions create short-run misconduct incentives, (iii) limited provider campaigns in rural areas, and (iv) misguided vendor beliefs about profit-maximizing prices[6]. We do not aim to separate the relative importance of the various ranked hypotheses, but the rankings are clear, robust, and preserved even with

[6] In section V, we explore (i) what explains the possibly misguided vendor beliefs and non-profit-maximizing prices and (ii) why the provider's information campaigns in rural areas are limited yet beneficial to the provider.

alternative scoring mechanisms. As shown, the issue of poorly-informed consumers, which in itself can exacerbate effects from the other possible hypotheses, is very crucial, and further demonstrates that information frictions matter. Details about the follow-up surveys with managers and vendors are contained in Appendix D.

III Experiment: Design

Intervention and Timetable. We evaluate the impacts on both customers and vendors of different information sets that reduce market misconduct. Our markets feature a version of the prisoner's dilemma problem: vendors (including other parties in the market) would be collectively better off if vendors did not cheat, but there are private benefits to deviating from a low-misconduct equilibrium, and they therefore end up in a privately profitable high-misconduct equilibrium. In such a market setting, with significant information frictions, it might be difficult to establish reputation and achieve the better collective outcome. Transparency and monitoring systems that enforce a low-misconduct equilibrium could be welfare improving, which we discuss below.

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

- Treatment program I: Price Transparency (PT) – Addresses the question of, "what to ask vendors while at vendor points." It informs consumers about the official tariffs for common local transactions, and thus improves consumer sophistication at detecting misconduct.

- Treatment program II: Monitoring and Reporting (MR) – Addresses the question of, "how to report seller misconduct." It provides customers with a toll-free number to report suspected misconduct to authorities, and thus raises the potential cost of misconduct to vendors if caught. Punishment for vendor misconduct ranges from losing business license, to provider warnings, and to customers not transacting at vendor points.

- Treatment program III: Combined PT+MR – A joint program that tests the interaction of programs I and II. See Exhibits in Appendix D for the specific information sets.

- Control program: no additional information.

We visit the assigned local markets three consecutive times over a two-month period (once every 2-3 weeks) to first deliver and then repeat the information programs to subjects. We conclude visits by asking subjects to summarize the information they received, and giving them hard copies of the treatment program. We ensure that vendors are equally aware of the interventions by communicating the same information to them, right after seeding the information with nearby households, yielding a two-sided information design. Together, our treatment programs aim to reduce potential information frictions and increase the social cost of vendor misconduct. Our design mitigates against Hawthorn effects, since all markets receive regular visits.

To roughly gauge the likely significance of the information programs, the recipients are ex-ante asked to rate the usefulness of the information we provide for their financial decision-making (customers) and for their businesses (vendors). We use a five-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.[7] Programs I (PT) and II

[7] In practice, our research team received around 75 different phone calls from the experimental subjects (specifically the customers) to discuss their M-Money two to three months after the provision of the information programs. This is a costly action (because consumers had to pay to call/ talk), represents about 9.3% of the treatment sample, and suggests that subjects are willing to pay for our information programs, perhaps because they find the information credible. In addition, this suggests

(MR) are popular consumer protection policy instruments in practice (Garz et al. 2021). By benchmarking the two programs against each other and against Program III (PT+MR), we can evaluate their relative effectiveness in reducing market misconduct committed against consumers, and assess whether Program I is compatible with Program II, or whether it only becomes effective when combined with an alternative that increases the direct cost of misconduct to firms. Table 1 shows the timetable of all field activities.

Data Collected. We gather information from multiple sources and rounds of data collection (Table 1): (i) combined listing and baseline market census (discussed earlier); (ii) baseline audit study (approach discussed below); (iii) transaction networks data; (iv) 22-weeks follow-up (phone) market survey, 33-weeks administrative audit study, and market-level transaction data from the largest service provider, which we call an endline. The official tariffs did not change between baseline and endline.

Administrative Audit Data. To objectively measure true misconduct, in the absence of existing credible data, we implement an audit study procedure where auditors (experimental customers) are given cash to make actual M-Money transactions at vendor points. The transactions (12 in total) span multiple, common transaction types: cash-in, cash-out, and account opening.[8] As mentioned, tariffs on transactions are ex-ante set by the providers. To mimic the local market context, and properly capture misconduct, we recruit and use local residents, who are trained to follow a consistent approach to interacting with vendors, including using uniform language, provided in a short and transparent transaction script (Appendix D contains details).

We randomly assign the local shoppers / auditors to a unique set of vendors and multiple transactions (the 12 different transaction types) are performed at random at each vendor

that subjects' rating of the usefulness of the information provided is less likely affected by potential experimenter demand (pleasing) effects (de Quidt, Haushofer, and Roth 2018).

[8]Importantly, these include transactions are regular in this marketplace and inherently mimic the nonlinear fee structure. The typical fee/ tariff structure set by providers is piecewise linear: GHS0.50 for all transaction values $\leq$GHS50, 1% of the value for transaction values between GHS50 and GHS1,000, and GHS10 for all transaction values $\geq$1,000. Similarly, the cost of a new SIM card is GHS2.0, and registering for a new M-Money wallet is free, but requires an initial minimum account deposit of GHS5.0. Appendix D and Table B.9 contain details.

point, as long as such services are available. There are instances where auditors are unable to make certain transactions for a variety of reasons, including unavailability of network and vendor's insufficient liquidity (*e*-credit or cash). With transaction-type fixed effects, as we do later in the empirical analysis of misconduct, such service interruptions have limited impact on our results. About four successful trips were made per auditor per day to their assigned vendors.

A potential concern with the audit measurement approach is that vendors cheat strangers (like the auditors) but not local repeat customers whom they know. This is not a major concern here, for several reasons. First, it might be more risky to cheat strangers, because they might be more informed, which is especially true in this market context with much imperfect information. This reduces the possibility that vendors systematically cheat strangers. Second, in our market environment, we estimate that a very large share of market transactions are conducted with customers who are not a family member or close relation of the vendor. Customers from our study area were shown the locality-level roster of all vendors and asked to indicate where they last transacted and how they are related to that vendor: 8.0% of transactions were between participants who are blood-related, 22.0% were between participants who are friends, and 70.0% were between unrelated participants. Third, we vary the type of transactions, and auditors conduct multiple or repeat transactions at a vendor point to mimic repeat customers. Auditors were, however, reassigned to different vendors between the baseline and endline to prevent vendors from identifying them at endline. We are confident that our audit-based measurement provides an unbiased estimate of the degree of misconduct.

We implement several quality controls for the transactional exercises. First, we set up a computer-adaptive data collection platform (called data HQ), which allows us to track and verify the data in real time and space. Right after every visit, auditors complete a brief questionnaire about the transaction using their tablets, out of sight of the vendors (see Table D.1), and synchronize the data to our data HQ for immediate access and verification. The

GPS coordinates of all transactions are traceable. Second, we piloted the proposed audit approach in February 2017 (as noted in the Market Census section), which yielded patterns of misconduct similar to the main experiment. Third, we include transaction types that are either easy or difficult for the seller to overcharge, finding consistent evidence of higher misconduct for the easy to overcharge transactions, as discussed below. Together, these quality controls strengthen our approach by 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).

We define misconduct to entail transactions that are over-charged when compared to the provider-approved tariff rates (as in Egan, Matvos, and Seru 2019; Annan 2020). Table B.9 and Figure B.6 show baseline results across the various transactions. We estimate that 22% of transactions are overcharged (which reflects the incidence of misconduct), which results in GHS3.3 (= 82% of the official tariffs) overpaid to the vendor (which reflects the severity of misconduct). There is heterogeneity in misconduct levels across the different types of transactions. Misconduct is concentrated in over-the-counter (OTC) transactions, which involve little to no automation or active verification from the customer, and are thus more vulnerable to vendor misconduct. Non-OTC transactions (e.g., opening a new account) are also overcharged, but at a much lower rate. This is reassuring, and alleviates several potential concerns, including that auditors might be over- or under- measuring misconduct.

Market Survey Data. We measure several repeated outcomes at different stages of the study. For vendors, we measure sales revenue by soliciting transaction records for their M-Money business and non-M-Money services (if the vendor operates a bundled store).

With customers, we restrict attention to four relevant outcomes: (i) adoption and usage of money services: we ask whether households use money services, and if so, the transaction amount involved per week; (ii) savings on M-Money: we ask whether households saved on their money wallets within the month; (iii) specific shock experiences (such as health, revenue, and household expenditures) and risk mitigation: we ask whether customers experienced unexpected shocks that they could not financially remedy, providing an objective proxy for insurance (Dupas and Robinson 2013; Breza and Chandrasekhar 2019); and (iv) poverty. Since our study focuses on M-Money in low-income and poor environments, we field questions that allow us to directly examine poverty. We adapt a recently developed measure of poverty, called the "Simple Poverty Scorecard", that is rigorous, inexpensive, simple, and transparent (for details, see Schreiner 2015).[9]

With these combined measurements, we gather data from both sides of the market, which allows us to cross-validate the accuracy of the records. For example, one will expect increases in household money transactions to (positively) correlate with increases in nearby vendor sales revenue, all else equal. See Appendix D for definitions of relevant select variables.

Treatment Assignment. We use a 2×2 factorial design, randomizing the 130 randomly selected markets (as defined below) into four experimental anti-misconduct programs: PT-alone (31 markets ≡ 31 select vendors × 272 nearby customers); MR-alone (32 markets ≡ 32 select vendors × 257 nearby customers); combined program (35 markets ≡ 35 select vendors × 276 nearby customers); and control program (32 markets ≡ 32 select vendors × 185 nearby customers). We stratify based on districts, and all misfits are resolved and randomly assigned. Figure B.2 displays the spatial distribution of the market-level treatment assignments. We identify distinct markets, which limits potential cross-market spillovers: (i) As displayed, most localities are spatially distinct and (ii) Consumers report not switching to use different vendors other than the nearby, local vendors.

Balance and Validity of Design. We discuss two different levels of balance. First, we focus our study on randomly-selected markets 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), each with

[9]We estimate an overall poverty rate of 10.0%, which is low but very close to the official poverty statistics of Ghana that report the rate in 2017 as 12.6% for the Eastern Region, where our study is based (GLSS 7 Report, p.19). The slight difference (underestimate) in poverty rates may be linked to one of the following reasons: (i) our poverty measure (Shreiner 2015), which is a shortcut, underestimates poverty, or (ii) our 5-house radius around a given vendor rule for household surveying captures relatively richer households, or (iii) simply, time trends, since GLSS 7's estimate was in 2017, while our estimate reflects 2019. In addition to poverty, we examine impacts on shock mitigation by households, which are alternative poverty-relevant outcomes.

surrounding customers or households (range=5-47, average=20.8). To maximize statistical power, we randomly select one vendor and their nearby customers per locality for our study. We call this combination (selected vendor × nearby households) a randomly-selected market. Sample representativeness requires that being a randomly-selected 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 randomly selected in the pre-intervention period. We consider a number of different relevant outcomes, and show that neither side of the market demonstrates any observable differences across the two groups. Tables B.1 and B.2 report the results, where we find no difference across those markets selected and those not selected.

Second, we base our treatment analysis on a comparison of randomly-selected 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., PT-alone, MR-alone, and combined 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 receives an information treatment, 0 otherwise. We consider the various treatments separately and together (i.e., pooled) for a number of different outcomes, and show that neither side of the market demonstrates any observable differences across the two groups. Tables B.3 and B.4 report the results, providing strong evidence in favor of balance, with no difference across subjects i (households or vendors) in assigned (treated) and non-assigned (control) markets.

Attrition. Table B.5 displays the breakdown of response rates and attrition between baseline and endline. To maximize response rates at endline, trained field officers conduct multiple phone calls (see Figure B.5) at different times of the day, varying either weekdays or weekends, combined with manual contact tracing for subjects with inactive phone numbers. For the survey-based measurements (customers and vendors), we record an overall attrition rate of 18%, which is low, given that the business of M-Money is subject to a high degree of migration and operator turnovers. Attrition is non-differential across arms (Tables B.6 and B.7 show tests for statistical significance by treatment arm). For our endline audit measurements, 129 out of the 130 randomly selected vendors were reached, implying an attrition rate of just 0.8%.[10] We evaluate and show robustness of the main results to attrition.

IV Experiment: Results

We present and discuss the treatment effects. Since all our treatments are about information provision, we report both the pooled (any treatment) and separate treatment effect(s) of the information sets. 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)[11] 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 primarily to increase precision and to control for potential confounds (if any). For the pooled effects, $\mathbf{I}_{vd}$ is a 0-1 indicator for whether a locality received

[10] The interventions did not lead to significant vendor exits from the local market (demonstrating limited adverse selection effects). Rather, they reduced vendor misconduct, which is consistent with moral hazard effects (similar to Klein, Lambertz, and Stahl 2016).

[11] We have one continuous outcome (consumers' weekly usage of services; Figure B.4) with zero values. To account for this, we report results using an inverse hyperbolic sine transformation `asinh`.

any of the information programs, and thus β captures the (pooled) 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 PT-alone, MR-alone, and combined information sets, respectively (i.e., $\beta = (\beta_1, \beta_2, \delta)'$). We report cluster-robust standard errors for outcomes with more than one observation per locality and heteroskedasticity-robust standard errors when there is one observation per locality.

For robustness checks, which we relegate to Appendix C, several alternative models and inference procedures are allowed. First, we report alternative standard errors, including the wild bootstrap cluster-t and randomization inference. Second, to address the potential issue of multiple testing, we adjust p-values for multiple testing across families of outcomes, following the procedure presented in Romano and Wolf (2005). Third, to evaluate potential attrition bias, we report Lee (2009) attrition bounds, Imbens and Manski (2004) confidence sets, and Behaghel et al. (2015) attrition bounds. Fourth, in alternative models, we choose $\mathbf{X}_{ivd}$ using post-double-selection LASSO (Belloni et al. 2014).

Treatment Effects on Seller Misconduct and Consumers' Beliefs (1)

Seller Misconduct: We ask whether the information programs reduce misconduct. Table 2 reports the pooled and separate treatment effects, and shows that the intervention meaningfully reduced vendor misconduct (measured using actual audit transactions). We estimate a pooled effect of -21 pp (-72% of control mean) for misconduct incidence and -GHS0.60 for misconduct amount (-78% of control mean). The effects are economically much larger for the combined and MR-alone programs, however, the differences across the programs are barely distinguishable statistically. These results strongly confirm that the information programs are indeed anti-misconduct, yielding economically very large, and statistically significant decreases in both incidence (occurrence) and intensity (shift in the distribution) of seller misconduct.

Consumers' Subjective Beliefs: We evaluate whether the information sets affected consumers' beliefs about vendor misconduct. Following Bursztyn, González, and Yanagizawa-

Drott (2020), we elicit perceptions about seller misconduct (or honest behavior, otherwise) by asking customers at endline agree or disagree with the statement: "In my view, M-Money vendors generally overcharge customer transactions at vendor points." To incentivize the beliefs elicitation, we also ask consumers: "What's your estimate of the % of others (all vendors and customers in this locality) that will Agree (Disagree, otherwise) with this statement?" In each local market, the respondent with the closest guess receives 10GHS (see Appendix D). For a third measure, we also ask whether customers believe they have experienced overcharges at vendor points, as in the baseline. The three subjective belief measures, which reflect consumer belief in vendors' trustworthiness, are significantly positively correlated (p-value = 0.000).

We ask if consumers' views about honest vendor behavior at endline shifted in the direction of the information treatments. Table 3 reports the treatment effects (Figure 1 provides graphical illustration). There is strong evidence that the intervention meaningfully increases consumers' perceptions of honest vendor behavior. We estimate a pooled effect of +7.0 pp (+30% of control mean; p-value = 0.095). The perceived effects appear to be much larger for the combined program (+13 pp = +30% of control mean; p-value = 0.022). The change in perceptions reflect the reality that treated consumers now have the technologies to enforce vendors' trustworthy behavior using the channels activated – social sanctions and/or punishment. The results robustly replicate across all three measures of beliefs.

We evaluate the accuracy of consumers' views, and whether they updated their beliefs as a result of the information sets. We estimate a regression of perception of misconduct (subjective) against the interactions of the treatment variables and the audit measure of misconduct (objective) to examine how the intervention causes consumer perceptions to more closely correlate with the audit measure of misconduct. Tables 4 shows the results. We estimate a pooled effect of +27 pp (+51% of control mean) increase in customers' ability to correctly guess misconduct behavior. The combined information program had economically larger effects. These results (i) provide evidence of updated consumer beliefs i.e., increased ability of customers to accurately evaluate vendor behavior, and (ii) show increased consumer sophistication. Treated customers are significantly better calibrated (+51%) about vendor behavior relative to the control group.

These results strongly confirm that the information programs do not only reduce seller misconduct / prices, but also meaningfully correct consumer misperceptions and improve knowledge about misconduct. We next evaluate the broader impacts on consumers and businesses.

Treatment Effects on Consumers' Use of M-Money and Savings (2)

Graphical Evidence: We provide graphical illustration of the treatment effects on consumers' use of M-Money. Figure 2 plots the empirical cumulative distributions of the `asinh` of total transaction amounts per week at endline by treatment status. There is strong visual evidence of positive effects of the information programs on consumers' transactions. This implies increased usage of M-Money financial services as a result of the information programs. The effects do not seem to be driven by specific parts of the distribution.

M-Money Usage and Savings: Table 5 reports the estimated effects on usage of services (or demand) and savings, respectively. There is increased transaction amount per week, with a treatment effect of about +45.8%. The probability of using the financial services increased (7.3 pp =+10.0% of control mean). The impacts are much larger for the combined program (+54.0% increased usage), compared to the individual information sets. The results are similar for savings likelihood on M-Money. There is evidence of an increased savings rate by 7.5 pp (=+12.1% of control mean). Customers are significantly more likely to save on M-Money, with much larger impacts for the combined program (13.1pp = +21.0% of control mean compared to the individual information sets). A Wald test rejects the null that the savings effect from the combined program is equal to effects from either the PT-alone (p-value=0.048) or MR-alone information set (p-value=0.066).

We combine all the usage and savings outcomes (via principal component analysis (PCA)), finding that the effects are consistently larger for the combined program (Tables C.5 and

C.6). This is followed by the MR-alone information set. These results indicate that the MR-alone and PT-alone programs are informationally complementary, and that PT-alone (a popular consumer protection instrument) may not be sufficient unless combined with random information assignment about MR.

Treatment Effects on Business Revenues (3)

Did market vendors experience an increase in sales revenue for M-Money? If the consumer records, and hence the estimated treatment effects, are accurate, then one might expect direct increases in M-Money business transactions (all else being equal). Table 6 reports the estimated impacts. We find evidence for a large positive impact on revenues for M-Money (+GHS437 = +54.9% of control mean).[12] Defining business exits (or deaths) as vendors that were unreachable and/or had inactive registered phone numbers during our endline phone surveys, we do not find evidence for an impact on exits from the local market [13] There is limited significant difference in treatment effects across the different information programs. Our evidence of significant impact on revenues/ intensive margin is consistent with Brown, Hossain, and Morgan (2010), who examined retail sellers on Yahoo and eBay, specifically in a market setting with low transaction tariffs.

Spillover Effects (4)

In principle, all the treatment effects on the non-price and non-beliefs outcomes are spillovers. However, the experimental design implies two specific spillover effects, which we emphasize below.

Misconduct for Untreated Businesses: We find significant spillover effects (Table 7): untreated vendors located in treated localities or markets reduce their misconduct (-21 pp

[12] To further explore the market-level effects, we solicit administrative data from the largest service provider about market transactions that originate from localities in our study area during the endline period. A limitation of this provider's dataset is that it does not cover all our experimental localities and hence does not provide much variation across the separate market-level treatment arms. However, pooling all the treatments, we find an overall increase in transaction activities in the treated markets relative to the control markets, which is qualitatively consistent with our treatment effects on consumer and vendor outcomes (results omitted and available upon request).

[13] This is consistent with the very low attrition rate (0.8%) of the audit transactional exercises that require physical visits to the vendor. Recall that we make repeated endline calls (see Figure B.5), varying the days and time of day. In our market environment, defining business exists as unreachable vendors seems relevant, because active vendor phone numbers are required for the business of M-Money to be in operation. However, it is possible that businesses could simply be switching their registered numbers, which seems unlikely: one can replace the vendor phone number at no cost if lost; obtaining a completely new vendor number is costly and entails more paperwork.

pooled effect). This broader impact on vendor misconduct is consistent with misconduct being contagious with externality effects, which is typical of vertical markets (Tirole 1988, Chapter 4). Good vendor reputation might be difficult to build pre-experiment due to externality effects of misconduct, when combined with imperfect information between vendors and consumers.

Revenues for Non M-Money Services: Table 8 reports the estimated impacts on stores that also offer non-M-Money services. Meaningful positive treatment effects are reported for non-M-Money business services (+55.4% of control mean). Total business sales, which adds the sales revenues from both M-Money and non-M-Money services, increase (+52.4% of control mean). The large positive impacts on non-M-Money transactions for bundled stores suggests positive spillover effects of the information programs on overall local market activities.[14]

Decreased vendor misconduct and increased demand for financial services are beneficial to consumers; increased sales revenue from M-Money's line of business is beneficial to service providers; but is the average vendor better or worse off? From the audit transaction data, we estimate an average effective price of about GHS1.70 per GHS100 transaction value for control vendors, versus GHS1.10 per GHS100 transaction volume for treated vendors. With a treatment effect of +GHS450 increase in M-Money services, the treatment increases M-Money sales revenue from about GHS800 to GHS1250. Vendors earn sales commission as profits, so the treatment changed their average profits from $\frac{1.70}{100} \times 800$=GHS13.60 to $\frac{1.10}{100} \times 1250$=GHS13.75. This implies that vendors are unaffected, which is consistent with the estimated elasticity of around 1.0. If we account for the additional improvements in vendors' non-M-Money services (the positive externalities from bundling), then the average vendor is better off.

[14] We do not have individual-level data to separately look at spillovers on untreated consumers (in addition to sales on vendors in treated localities). Aggregate market-level data from the largest service provider, however, show increased transaction activities overall in the treated markets relative to the control markets. This implies a potentially positive spillover effect on usage of services for untreated consumers/ vendors, which is also consistent with the negative spillover effects on misconduct for untreated vendors who served untreated customers at baseline.

This market operates with a vertical structure: service providers at the upstream set up vendors at the downstream who work for them and earn commissions on sales, which leads to a version of the well-known single-versus-double marginalization problem with vendor misconduct (Tirole 1988, Chapter 4; Janssen and Shelegia 2015). Providers have limited visibility into the behavior of their vendors. Because service providers fix prices of transactions at vendor points (price forcing to charge marginal cost), vendors cannot reduce their sales in an attempt to marginalize. Through misconduct, vendors impose illegal mark-ups on transactions, which results in lower sales revenue than is optimal from the viewpoint of the provider. The treatment pushed the double-marginalization high price (high misconduct) to a single-marginalization lower price (low to no misconduct). In this case, lowered misconduct results in benefits not only to providers and vendors, but also to consumers. In our setting, there are vendors who earn profits not only from the M-Money business, but also from selling other products. When the treatment leads to less misconduct, customers conduct larger money transactions and also purchase other non-M-Money items at the vendors premises. Thus, we show that the vendor can also be better off under the single marginalization result, which is a novel and interesting result.[15]

Treatment Effects on Shocks Mitigation (5)

Mitigation of Shocks: Revenue, Health, and Expenditure: Did customers (or households) increase their transactional services and savings likelihood in meaningful enough levels that they are better able to mitigate unexpected shocks? Tables 9 and 10 show the estimated effects on customers' experiences of unmitigated shocks and poverty. We report on general shocks (any experience), and, individually, on shocks related to household revenue, health, and household expenditures.

There is reduced instance(s) of general unexpected shocks that consumers could not financially remedy or pay for (i.e., when resource limits bind) (-6.8 pp =-7.6% of control mean, p-value=0.044). This effect is mainly driven by household expenditures, which has

[15] We thank Matt Shum for pointing this out. We note, however, that we do not have a direct counterfactual where both the upstream provider and the regulator allow double marginalization. We have an approximation that uses the fact that, by committing misconduct, the downstream vendors impose illegal markups (in addition to the markup the upstream provider has imposed). There are two indirect counterfactuals: (i) control vendors committing misconduct and (ii) treatment vendors committing misconduct in the *pre*-experiment phase.

the largest significant reduction of 10.4 pp (p-value=0.080). However, bot revenue and health sources are equally meaningful based on their effect sizes (7.2 pp and 6.1 pp, respectively). For shock mitigation, the PT-alone and combined information programs show significantly negative larger impacts. Effects from the MR-alone program are relatively small and insignificant. These estimates provide a large and objective proxy for financial resilience and insurance value of reducing seller misconduct to consumer welfare. We do not find evidence for an impact on overall poverty levels.

V Discussions and Interpreting the Results

A Discussions

The broader improvements in consumer and business outcomes are interesting; however, these raise three immediate questions and/ or implications.

Implication 1: Why do vendors overcharge — i.e. set higher prices — that do not necessarily maximize profits but decrease total welfare? To explore this question, in a follow-up exercise at endline, we solicit the beliefs of vendors in control markets (n=58) about prices and then ask the vendors to predict the intervention's likely effect on prices and quantities (treatment effects) [à la DellaVigna and Pope 2018]. The exercise suggests that vendors commit (unprofitable) misconduct because they perceive that a higher price is better than a lower price. In our context, such perception is reasonable, because while vendors can predict very well prices, which they set, they cannot predict well the effect on quantities following a price change (i.e., they cannot predict well the price elasticity), leading them to put less weight on quantity effects. In short, retailers seem to overcharge because of their inability to predict well the price elasticity of demand. Details are contained in Appendix C.

Implication 2: Why did the provider not implement interventions similar to our proposed two-sided information programs, despite the promise of improving provider revenues? To explore this question, managers – working for the provider (n=29) – were invited to predict effects of the information interventions on prices and quantities (treatment effects) [à la DellaVigna and Pope 2018]. Details are contained in Appendix C. Most managers were systematically incorrect in their forecasts — predicting zero absolute effects for the interventions. We find suggestive evidence that managers (i) were unaware that such interventions and their specific informational contents will work and (ii) perceive the cost of information campaigns in rural areas to be more expensive.

Implication 3: Benchmarking the Magnitude of Treatment Effects—The program impacts on quantities are very large. For context, the typical transaction is about GHS100 (based on the audit transactions of GHS50, GHS160 and GHS1100, which were chosen to be typical of the market setting, Table B.9). The official fee will be 1% of this transaction value, which implies a fee value of GHS1.0. The experiment leads the total fee (official fee + misconduct) to fall from about 1.70% to about 1.10% (Table 2), about a 40% reduction of the transaction fee. The 45% increase in consumer demand (or 52% increase in vendor's total sales revenue) in response to a 40% fee reduction is reasonable; it is an elasticity of about 1.1 (or 1.3). Our estimated impacts are very reasonable, and fall within the range of market effects from relevant M-Money tax and subsidy policy experiments (see details in Appendix C).

B. Heterogeneity

We examine heterogeneity along five dimensions: (i) vendor competition (market conditions), (ii) seller's gender (market conditions), (iii) pre-experiment consumer illiteracy (compliers of the information programs), (iv) bundled stores, and (v) beliefs update effect on consumer outcomes (compliers of the information programs). The results and details are contained in Appendix C. The reduction in misconduct is concentrated in localities with more competition and a high fraction of uninformed customers, as well as within vendors who bundle services, but the effects are similar across gender. In addition, the downstream effects on quantities are concentrated in localities where consumers significantly updated their beliefs about vendor misconduct.

C. Interpreting the Results – A Descriptive Model of Reputation

We seek to understand what happens when we give relevant seller misconduct information to both (potentially uninformed) consumers and (potentially dishonest and informed) vendors in a local finance. One could tell several stories about how the information intervention might act to affect misconduct and, thus, market outcomes. Our underlying hypothesis, however, is that vendors in our experiment expect they are more likely to be perceived by potential customers as irresponsible if they commit misconduct. Following Macchiavello and Morjaria (2015), we define a vendor's reputation as consumer perceptions about the vendor's tendency to commit misconduct (vendor's dis/honesty). In Appendix A, we present a model that formalizes these arguments.

To organize ideas from the model for our reputation interpretation, it is useful to define some terms while mapping the model to our experiment. Market vendors decide whether to commit misconduct ($s = 0$) or not ($s = 1$). Customers imperfectly observe the vendor's action s. Denote by $\hat{\pi}$ consumers' imperfect belief about the probability that a vendor is honest. Consumers (uninformed vs informed) learn about the transactional action through public signals σ, giving rise to a moral hazard problem (Board and Meyer-ter-Vehn 2013). Based on their inference about a vendor's action given the available signal, a customer either assigns a reputational payoff ($\mathbb{E}[\hat{s} = 1|\sigma]$) to the vendor (via either PT or MR information programs) or reports the vendor's dishonest behavior as a direct punishment (via MR information program). If customers perceive (via $\hat{\pi}$) that the vendor is honest, then the vendor receives higher revenue (i.e., through repeated or large transactions and not being reported).

Our central goal is to compare market-level information sets about misconduct: one "without" information and another "with" information assignment about anti-misconduct. For the information assignment, we vary the information sets: one with technology to detect and reward misconduct behavior (reputation effects, where $\sigma = s$), another with technology to directly report and punish misconduct behavior (reputation and punishment effects), and one with both. We model assignment of the anti-misconduct market information as either a shift in the distribution of $\hat{\pi}$ or $\mathbb{E}[\hat{s} = 1|\sigma]$, which measures reputational concerns for sellers.

We document three pieces of evidence based on the model to show that reputation is at play: (i) changes in consumer beliefs about misconduct, $\hat{\pi}$, which we measure as the impact of the information treatment on $\hat{\pi}$ (Table 3); (ii) updates in consumer beliefs, $\mathbb{E}[\hat{s} = 1|\sigma]$, which we measure as the impact of the information treatment on the correlation between $\hat{\pi}$ and the audit measure of misconduct (Table 4); and (iii) the treatment effects on quantities are concentrated in markets where consumers significantly updated their beliefs about vendor behavior (compliers of the information programs) (Table C.15). The two-sided information programs attenuate consumer misbeliefs about misconduct and raise vendor concern for reputation.

D. Alternative Interpretations

The information programs substantially improved consumers' ability to infer misconduct accurately, and to report it, while also increasing vendor reputation concerns. As a result, prices/ misconduct decreased and quantities / consumer demand, including firm revenues and shock-mitigation increased. However, the information interventions might also turn on at least five other interesting alternative interpretations of the results: (i) Hawthorne effects, (ii) selection effects, (iii) marketing effects, (iv) price and/or (v) bargaining effects. In Appendix C, we find limited support for alternative explanations. We further estimate that about 55% of our information treatment effect is attributable to bias effect from consumers misperceiving seller misconduct, versus 45% for price effect from price increases as a result of seller misconduct.

E. The Value of Anti-Misconduct Information

To put our treatment effects into context, we consider the cost-effectiveness of providing anti-misconduct information to local markets. We use a very conservative approach that focuses on a usage of money services-only measure for customers and a sales revenue-only measure for vendors, ignoring the broader positive impacts. We estimate a very large value for the anti-misconduct information sets: (i) a cost-effectiveness ratio of 1:5 for consumers, implying a per subject cost of US$4.0 for about +US$19.3 increase in usage of services; and (ii) for vendors, a ratio of 1:21 improvement in revenue. These rough estimates compare favorably with other financial information programs (Frisancho 2018; Kaiser et al. 2020).

Details are in Appendix C.

VI Conclusion

Misconduct in markets with imperfect information matters for efficiency. The provision of information sets that deter and reduce retail vendor misconduct has broader market impacts. Customers meaningfully increase their demand for transactional services and their savings behavior at vendor points, which enables them to better mitigate unexpected shocks. Businesses experience meaningful increases in their sales revenue, with limited impact on vendor profits/ commissions, suggesting improved market efficiency.

Reputation does matter for misconduct. In rural financial environments, where markets are subject to a high degree of imperfect information, the use of reputation as a discipline device against market misconduct is limited. When customers do not know official and mandated prices, they cannot observe whether they are being cheated, making it difficult for vendors to establish a good reputation—which may increase vendor misconduct. However, reputation becomes an effective tool and disciplinary if there is a high probability that customers will infer misconduct (Shapiro 1982, Burkhardt 2018), and if vendors can easily demonstrate the quality of their market services. Such reputation-driven conduct is illuminated drawing on features of our empirical setting and the provision of relevant market-level information that improves subjects' ability to report misconduct and make accurate inferences about it.

Our field experiment is carefully designed to (i) reduce market vendor misconduct/ prices through cost-effective information programs; (ii) quantify the programs' impact on important economic outcomes/ quantities on both sides of the market; and (iii) show that these effects are driven by a combination of more accurate consumer beliefs about misconduct and increased vendor concern for reputation. 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 are underprovided (Jensen and Miller 2018; Bai 2019) due to seller misconduct, which diminishes overall market efficiency.

Commerce requires reputation and/ or consumer trust, but reputation in markets might be difficult to build, and thus low, due to imperfect information. In developing countries, where consumers are either uninformed about FinTech or lack experience with it, and many market digitization initiatives are underway, consumers suffer significant market misconduct, which can lead to consumer mistrust in payment markets, act as a barrier to market activities, and reduce households' welfare. Our study is the first, to our knowledge, to provide quantitative estimates on vendor misconduct and the value of anti-misconduct information programs in payment markets, emphasizing the effect of social sanctions and punishment.

References

[1] Annan, Francis. 2017. "Fraud on Mobile Financial Markets: Evidence from a Pilot Audit Study." NET Institute - Working Paper No. 17-16.

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

[3] Atkin, David, Amit K. Khandelwal, and Adam Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment." Quarterly Journal of Economics, 132 (2): 551-615.

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

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

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

[7] Behaghel, Luc, Bruno Crepon, Marc Gurgand, and Thomas Le Barbanchon. 2015. "Please Call Again: Correcting Non-Response Bias in Treatment Effect Models." Review of Economics and Statistics, 97 (5): 1070-1080.

[8] 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.

[9] Blackmon, William, Rafe Mazer and Shana Warren. 2021, "IPA Consumer Protection Research Initiative: RFP Overview."

[10] Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." Quarterly Journal of Economics, 128 (1): 1-51.

[11] BMGF. 2021. "Research Brief: The Impact of Mobile Money on Poverty." Bill and Melinda Gates Foundation.

[12] Board, Simon and Moritz Meyer-ter-Vehn. 2013. "Reputation for Quality." Econometrica, 81 (6), 2381-2462.

[13] Breza, Emily L. and Arun Chandrasekhar. 2019. "Social Networks, Reputation, and Commitment: Evidence from a Savings Monitors Experiment." Econometrica, 87 (1): 175-216.

[14] Brown, Jennifer, Tanjim Hossain, and John Morgan. 2010. "Shrouded Attributes and Information Suppression: Evidence from the Field." Quarterly Journal of Economics, 125 (2): 859-876.

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

[16] Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott. 2020. "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." American Economic Review, 110 (10): 2997-3029.

[17] Cai, Jing and Adam Szeidl. 2017. "Interfirm Relationships and Business Performance." Quarterly Journal of Economics, 133 (3): 1229-1282.

[18] DeLiema, Marguerite, Martha Deevy, Annamaria Lusardi and Olivia S. Mitchell. 2018. "Financial Fraud Among Older Americans: Evidence and Implications." NBER Working Paper.

[19] DellaVigna, Stefano and Devin Pope. 2018. "Predicting Experimental Results: Who Knows What?", Journal of Political Economy, 126(6): 2410-2456.

[20] De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." Quarterly Journal of Economics, 123 (4): 1329-1372.

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

[22] Dupas, Pascaline and Jonathan Robinson. 2013. "Why Don't the Poor Save More? Evidence from Health Savings Experiments." American Economic Review, 103 (4): 1138-1171.

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

[24] Egan, Mark, Gregor Matvos, and Amit Seru. 2022. "When Harry fired Sally: The Double Standard in Punishing Misconduct." Journal of Political Economy, 130 (5): 1184-1248.

[25] FORTUNE. 2002. Quoted in Clifton Leaf, "White-Collar Criminals: They Lie, They Cheat, They Steal, and They Have Been Getting Away With It for Too Long," Fortune (March 18, 2002), p. 62. Available here: https://money.cnn.com/magazines/fortune/fortune_archive/2002/03/18/319921/

[26] Garz, Seth, Xavier Giné, Dean Karlan, Rafe Mazer, Caitlin Sanford, and Jonathan Zinman. 2021. "Consumer Protection for Financial Inclusion in Low and Middle Income Countries: Bridging Regulator and Academic Perspectives." Annual Review of Financial Economics, 13(1): 219-246.

[27] 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.

[28] Gibbons, Robert S. and John Roberts. 2012. "The Handbook of Organizational Economics." Princeton University Press.

[29] GSMA. 2018. "The GSMA Mobile Money Certification." GSMA Association. Available here: https://www.gsma.com/mobilefordevelopment/resources/the-gsma-mobile-money-certification/

[30] GSMA. 2020. "The Causes and Consequences of Mobile Money Taxation: An Examination of Mobile Money Transaction Taxes in sub-Saharan Africa." GSMA Association.

[31] Hasanain, Syed A., Muhammad Yasir Khan, and Arman Rezaee. 2023. "No Bulls: Experimental Evidence on the Impact of Veterinarian Ratings in Pakistan." Journal of Development Economics, 161(102999).

[32] Higgins, Sean. 2020. "Financial Technology Adoption." Conditionally Accepted, American Economic Review.

[33] Imbens, Guido and Charles Manski. 2004. "Confidence Intervals for Partially Identified Parameters." Econometrica, 72 (6): 1845-1857.

[34] Jack, William, and Tavneet Suri. 2014. "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." American Economic Review, 104 (1): 183-223.

[35] Janssen, Maarten, and Sandro Shelegia. 2015. "Consumer Search and Double Marginalization." American Economic Review, 105 (6): 1683-1710.

[36] 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.

[37] Lee, David. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." Review of Economics Studies, 76 (3): 1071-1102.

[38] Kaiser, Tim, Annamaria Lusardi, Lukas Menkhoff, and Carly J. Urban. 2020. "Financial Education Affects Financial Knowledge and Downstream Behaviors." NBER Working Paper No. 27057.

[39] Karpoff, Jonathan M. 2012. "Does Reputation Work to Discipline Corporate Misconduct?" Ch. 18, pages 361-382 of: Oxford Handbook of Corporate Reputation.

[40] 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.

[41] Klein, Tobias J., Christian Lambertz, and Konrad O. Stahl. 2016. "Market Transparency, Adverse Selection, and Moral Hazard." Journal of Political Economy, 124 (6): 1677-1713.

[42] Macchiavello, Rocco and Ameet Morjaria. 2015. "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports." American Economic Review, 105 (9): 2911-2945.

[43] Matsa, David A. 2011. "Competition and Product Quality in the Supermarket Industry." Quarterly Journal of Economics, 126 (3): 1539-1591.

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

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

[46] Romano, Joseph. P. and Michael Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” Econometrica 73(4): 1237-1282.

[47] Rosenthal, Robert W. 1980. “A Model in Which an Increase in the Number of Sellers Leads to a Higher Price.” Econometrica, 48 (6): 1575-1579.

[48] Satterthwaite, Mark A. 1979. “Consumer Information, Equilibrium Industry Price, and the Number of Sellers.” Bell Journal of Economics, 10 (2): 483-502.

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

[50] Shapiro, Carl. 1982. “Consumer information, Product Quality, and Seller reputation.” The Bell Journal of Economics, 13 (1): 20-35.

[51] Shapiro, Carl. 1983. “Premiums for High Quality Products as Returns to Reputations.” Quarterly Journal of Economics, XCVIII: 658-679.

[52] Suri, Tavneet and William Jack. 2016. “The Long-run Poverty and Gender Impacts of Mobile Money.” Science, 354 (6317): 1288-1292.

[53] Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

[54] TCI. 2023. “Transaction Cost Index: Year 1 Comparative Report”. Innovations for Poverty Action (IPA).

[55] Zinman, Jonathan and Eric Zitzewitz. 2016. “Wintertime for Deceptive Advertising?”, American Economic Journal: Applied Economics, 8 (1): 177-192.

[56] Zitzewitz, Eric. 2012. “Forensic Economics.” Journal of Economic Literature, 50 (3): 731-769.

Main Results for Text

Table 1: **Study timeline**

	DATE	ACTIVITY
Part 1	February 2017	**Pilot:** Misconduct – Incidence, severity, and correlates (Annan 2017)
Part 2	Feb 15-Mar 20, 2019	**Baseline:** Market census
	Aug 1- Aug 15, 2019	**Select sample (Experiment)** **Intervention:** Information sets assignment
	Sep 01-Oct 15, 2019	**Audit study I (Baseline)**
Part 3	Oct 15- Dec 15, 2019	**Intervention:** Information sets deployment
Part 4	May 15-May 30, 2020	**Endline:** Phone surveys + manual tracing supplement
	Aug 15-Sep 01, 2020	**Audit study II (Endline)**
	> Sep 15, 2020	**Administrative data:** Market transaction records from service provider

Table 2: **Prices: effect of information sets on vendor misconduct**

	Misconduct indicator		Misconduct amount (GHS)	
	(1)	(2)	(3)	(4)
Any treatment (β)	-0.211		-0.551	
	(0.086)		(0.255)	
	[-0.382, -0.039]		[-1.059, -0.042]	
Transparency alone (β_1)		-0.184		-0.439
		(0.094)		(0.276)
		[-0.372, -0.003]		[-0.988, 0.110]
Monitoring alone (β_2)		-0.217		-0.574
		(0.093)		(0.275)
		[-0.403, -0.030]		[-1.122, -0.027]
Combined (δ)		-0.212		-0.554
		(0.089)		(0.279)
		[-0.390, -0.033]		[-1.110, -0.001]
Observations	335	335	335	335
Mean of dep var in control	0.294	0.294	0.778	0.778
p-value (test: $\beta_1 = \delta$)		0.670		0.553
p-value (test: $\beta_2 = \delta$)		0.921		0.923
p-value (test: $\beta_1 = \beta_2$)		0.563		0.411
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.108		0.204

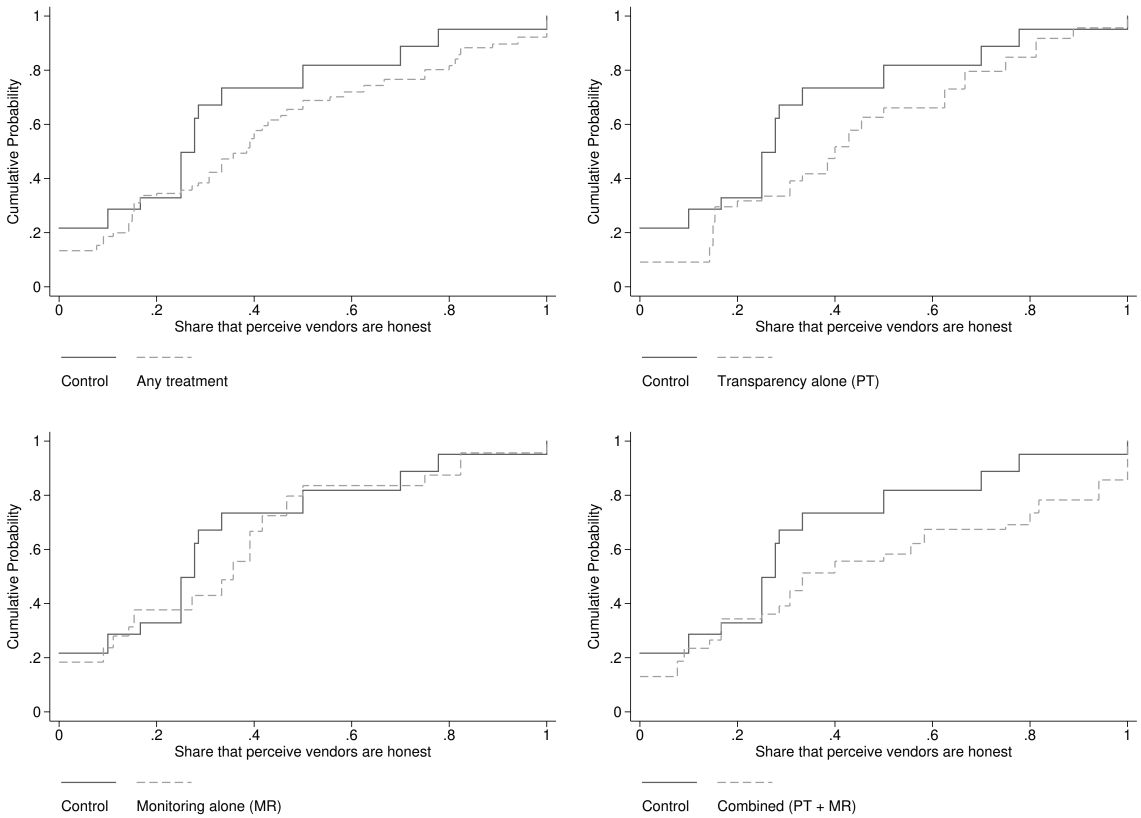
Note: Observations are at the select vendor x transaction type x transaction date level. Dependent variables are audit-based measures. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Table 3: **Consumers' beliefs about vendor honesty increases**

	Belief about vendor honesty indicator	
	(1)	(2)
Any treatment (β)	0.070	
	(0.040)	
	[-0.011, 0.145]	
Transparency alone (β_1)		0.107
		(0.057)
		[0.007, 0.221]
Monitoring alone (β_2)		-0.045
		(0.057)
		[-0.158, 0.068]
Combined (δ)		0.126
		(0.054)
		[0.018, 0.233]
Observations	810	810
Mean of dep var in control	0.314	0.314
p-value (test: $\beta_1 = \delta$)		0.747
p-value (test: $\beta_2 = \delta$)		0.005
p-value (test: $\beta_1 = \beta_2$)		0.022
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.432

Note: Observations are at the customer level. Dependent variable is a survey-based measure. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Belief denotes customers' perception that they are not being overcharged at vendor points (or customers' perception that they have not experienced seller misconduct) at endline. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Figure 1: **Consumers' beliefs about vendor honesty impacts by treatment**



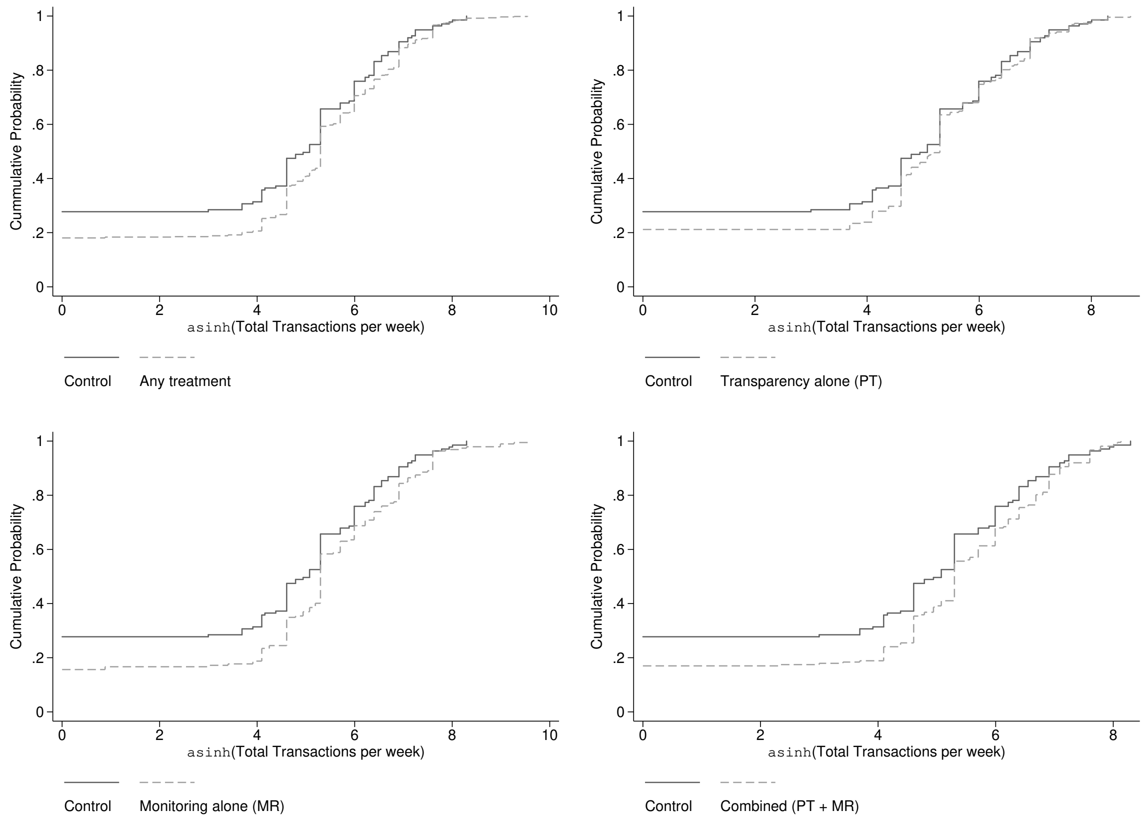
Note: Figure plots distributions (CDFs) of customer beliefs about honest vendor behavior at endline for the different experimental subsamples. Observations are at the customer level. Belief denotes customers' perception that they are not being overcharged at vendor points (or perception that they have not experienced seller misconduct). In each local market, we compute the share of experimental customers who indicate they believe they are not experiencing misconduct (indicating belief in honest vendor behavior) at endline. From a Kolmogorov–Smirnov test for the equality of distributions, p-value $= 0.000$ for all cases.

Table 4: **Belief update: effect of information sets on correct inference about vendor misconduct**

	Belief about vendor misconduct indicator	
	(1)	(2)
Any treatment (β)	-0.282	
	(0.082)	
	[-0.445, -0.119]	
x Objective misconduct	0.273	
	(0.106)	
	[0.062, 0.483]	
Transparency alone (β_1)		-0.365
		(0.087)
		[-0.537, -0.192]
x Objective misconduct (b_1)		0.349
		(0.122)
		[0.107, 0.592]
Monitoring alone (β_2)		-0.152
		(0.093)
		[-0.338, 0.033]
x Objective misconduct (b_2)		0.235
		(0.121)
		[-0.004, 0.475]
Combined (δ)		-0.354
		(0.078)
		[-0.510, -0.199]
x Objective misconduct (d)		0.284
		(0.109)
		[0.067, 0.501]
Objective misconduct	-0.199	-0.216
	(0.087)	(0.082)
	[-0.373, -0.0255]	[-0.380, -0.053]
Observations	810	810
Mean of dep var in control	0.685	0.685
p-value (test: $b_1 = d$)		0.586
p-value (test: $b_2 = d$)		0.683
p-value (test: $b_1 = b_2$)		0.385
p-value (test:$b_1 + b_2 = d$)		0.082

Note: Observations are at the customer level. Dependent variable is a survey-based measure. Objective misconduct is an audit-based measure. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Belief denotes customers' perception that they are being overcharged at vendor points (or perception that they have experienced seller misconduct) at endline. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Figure 2: **Quantities: consumer transactions impacts by treatment**



Note: Figure plots the distributions (CDFs) of `asinh` (Total transactions per week) at endline for the different experimental subsamples. Observations are at the customer level. From a Kolmogorov–Smirnov (KS) test for the equality of distributions, p-values equal 0.091, 0.481, 0.068 and 0.065, respectively (for equality tests, we trimmed the consumer transactions data at the 5% level). Equality tests reject the null that the distributional pairs are equal in all cases (p-values$<$0.091) except for the PT-only program (p-value=0.481).

Table 5: **Quantities: effect of information sets on usage of services and savings**

	asinh (Total transactions per week)		Used M-Money (last month)		Saved (last month)	
	(1)	(2)	(3)	(4)	(5)	(6)
Any treatment (β)	0.458		0.073		0.075	
	(0.225)		(0.039)		(0.042)	
	[0.011, 0.905]		[0.004, 0.151]		[0.008, 0.158]	
Transparency alone (β_1)		0.262		0.048		0.047
		(0.263)		(0.044)		(0.047)
		[-0.260, 0.784]		[-0.040, 0.137]		[-0.046, 0.141]
Monitoring alone (β_2)		0.587		0.084		0.042
		(0.268)		(0.044)		(0.052)
		[0.056, 1.119]		[-0.004, 0.172]		[-0.061, 0.146]
Combined (δ)		0.540		0.087		0.131
		(0.255)		(0.043)		(0.048)
		[0.035, 1.046]		[0.001, 0.174]		[0.035, 0.227]
Observations	810	810	810	810	810	810
Mean of dep var in control	4.096	4.096	0.734	0.734	0.622	0.622
p-value (test: $\beta_1 = \delta$)		0.215		0.248		0.048
p-value (test: $\beta_2 = \delta$)		0.832		0.917		0.066
p-value (test: $\beta_1 = \beta_2$)		0.183		0.319		0.919
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.367		0.420		0.536

Note: Observations are at the customer level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Total transactions per week is the value of M-Money transactions customer conducted in the local market per week at endline. Used M-Money (last month) is a 0-1 indicator for whether the customer used M-Money at endline. Saved (last month) is a 0-1 indicator for whether the customer saved money on M-Money at endline. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Table 6: **Effect of information sets on mobile money revenue and business exit**

	Sales (M-Money) per day (GHS)		Business exit indicator	
	(1)	(2)	(3)	(3)
Any treatment (β)	436.6		-0.069	
	(178.4)		(0.058)	
	[82.12, 791.1]		[-0.184, 0.046]	
Transparency alone (β_1)		523.6		-0.100
		(222.0)		(0.060)
		[82.44, 964.8]		[-0.220, 0.020]
Monitoring alone (β_2)		418.4		-0.094
		(259.8)		(0.063)
		[-96.93, 934.8]		[-0.220, 0.032]
Combined (δ)		358.1		-0.017
		(198.1)		(0.076)
		[-32.55, 751.8]		[-0.168, 0.132]
Observations	107	107	129	129
Mean of dep var in control	792.8	792.8	0.218	0.218
p-value (test: $\beta_1 = \delta$)		0.436		0.200
p-value (test: $\beta_2 = \delta$)		0.810		0.213
p-value (test: $\beta_1 = \beta_2$)		0.680		0.888
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.096		0.053

Note: Observations are at the select vendor level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Business exits (or deaths) are defined as vendors that were unreachable and/or had inactive registered phone numbers during our endline phone surveys. Heteroskedasticity-robust standard errors are reported in parenthesis. 95% confidence intervals are reported in brackets.

Table 7: **Spillover effects of information sets**

	I. MISCONDUCT FOR UNTREATED BUSINESSES			
	Misconduct indicator		Misconduct amount (GHS)	
	(1)	(2)	(3)	(4)
Any treatment (β)	-0.218		-0.648	
	(0.065)		(0.206)	
	[-0.348, -0.088]		[-1.060, -0.235]	
Transparency alone (β_1)		-0.232		-0.720
		(0.070)		(0.196)
		[-0.374 -0.091]		[-1.113, -0.327]
Monitoring alone (β_2)		-0.239		-0.693
		(0.075)		(0.242)
		[-0.389, -0.089]		[-1.178, -0.207]
Combined (δ)		-0.178		-0.524
		(0.070)		(0.224)
		[-0.319, -0.037]		[-0.974 -0.075]
Observations	411	411	411	411
Mean of dep var in control	0.278	0.278	0.783	0.783
p-value (test: $\beta_1 = \delta$)		0.315		0.179
p-value (test: $\beta_2 = \delta$)		0.235		0.323
p-value (test: $\beta_1 = \beta_2$)		0.915		0.859
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.001		0.001

Note: Observations are at non-select vendor x transaction type x transaction date level. Dependent variables are audit-based measures. Estimations compare non-treated vendors located in treated localities to the pure control localities. Includes randomization strata (district) x transaction type x transaction date dummies. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Table 8: **Spillover effects of information sets**

	II. REVENUE FOR NON-MOBILE MONEY			
	Sales (Non M-Money) per day (GHS)		Sales (Total) per day (GHS)	
	(1)	(2)	(3)	(4)
Any treatment (β)	132.7		537.6	
	(58.67)		(195.8)	
	[16.19, 249.3.1]		[148.5, 926.7]	
Transparency alone (β_1)		167.1		733.8
		(73.40)		(249.1)
		[21.31, 313.0]		[238.7, 1228]
Monitoring alone (β_2)		80.51		448.2
		(65.86)		(279.0)
		[-50.34, 211.3]		[-106.3, 1002]
Combined (δ)		141.4		402.5
		(76.43)		(215.7)
		[-10.47, 293.2]		[-26.71, 830.7]
Observations	107	107	107	107
Mean of dep var in control	239.5	239.5	1032	1032
p-value (test: $\beta_1 = \delta$)		0.748		0.173
p-value (test: $\beta_2 = \delta$)		0.330		0.862
p-value (test: $\beta_1 = \beta_2$)		0.223		0.306
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.270		0.043

Note: Observations are at the select vendor level. Dependent variables are survey-based measures. Includes (i) randomization strata (market district) dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Heteroskedasticity-robust standard errors are reported in parenthesis. 95% confidence intervals are reported in brackets. About 75% of vendors bundled M-Money with other business services. Non-mobile money sales code to zero for all outlets that only provide mobile money services. Total sales per day combine mobile money and non-mobile money sales revenues.

Table 9: **Effect of information sets on shock mitigation and poverty**

	u-shocks experience (any)		Poverty likelihood (%)	
	(1)	(2)	(3)	(4)
Any treatment (β)	-0.068		1.161	
	(0.033)		(1.355)	
	[-0.135, -0.002]		[-1.521, 3.844]	
Transparency alone (β_1)		-0.099		1.755
		(0.039)		(1.630)
		[-0.176, -0.021]		[-1.472, 4.983]
Monitoring alone (β_2)		-0.015		1.824
		(0.038)		(1.522)
		[-0.091, 0.061]		[-1.189, 4.838]
Combined (δ)		-0.085		0.001
		(0.045)		(1.640)
		[-0.174, -0.003]		[-3.245, 3.247]
Observations	810	810	810	810
Mean of dep var in control	0.895	0.895	9.899	9.899
p-value (test: $\beta_1 = \delta$)		0757		0.262
p-value (test: $\beta_2 = \delta$)		0.104		0.209
p-value (test: $\beta_1 = \beta_2$)		0.027		0.959
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.618		0.107

Note: u denotes unmitigated and is a 0-1 indicator for whether consumer experienced unexpected shock(s) that s/he could not financially remedy or pay for. Observations are at the customer level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Poverty is in percent (%) and measured using the Simple Poverty Scorecard (for details, see Schreiner 2015). Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Table 10: **Effect of information sets on shock mitigation**

	u-Shocks HH revenue		u-Shocks health		u-Shocks HH expenditure	
	(1)	(2)	(3)	(4)	(5)	(6)
Any treatment (β)	-0.072		-0.061		-0.104	
	(0.048)		(0.061)		(0.059)	
	[-0.168, 0.024]		[-0.183, 0.061]		[-0.221, 0.012]	
Transparency alone (β_1)		-0.114		-0.093		-0.137
		(0.057)		(0.069)		(0.067)
		[-0.229, 0.001]		[-0.231, 0.045]		[-0.271, -0.003]
Monitoring alone (β_2)		-0.002		0.008		-0.035
		(0.057)		(0.072)		(0.075)
		[-0.115, 0.111]		[-0.134, 0.151]		[-0.186, 0.114]
Combined (δ)		-0.091		-0.090		-0.132
		(0.057)		(0.068)		(0.064)
		[-0.205, 0.022]		[-0.226, 0.044]		[-0.260, -0.004]
Observations	810	810	810	810	810	810
Mean of dep var in control	0.783	0.783	0.531	0.531	0.419	0.419
p-value (test: $\beta_1 = \delta$)		0.658		0.967		0.930
p-value (test: $\beta_2 = \delta$)		0.090		0.090		0.145
p-value (test: $\beta_1 = \beta_2$)		0.041		0.107		0.149
p-value (test: $\beta_1 + \beta_2 = \delta$)		0.738		0.945		0.665

Note: u denotes unmitigated and is a 0-1 indicator for whether consumer experienced unexpected shock(s) that s/he could not financially remedy or pay for. Observations are at the customer level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

Supplementary Appendix (For Online Publication)

A 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 seller misconduct information to both (potentially dishonest and informed) vendors and (potentially uninformed) consumers in a local finance context. One could tell several stories about how the information intervention might act to affect misconduct and thus 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 misconduct in our experiment. Following Macchiavello and Morjaria (2015), we think of a vendor's reputation as consumer perceptions about the vendor's tendency to commit misconduct. Negative perceptions trigger direct punishments and affect vendor reputation (via a reduction of vendor sales in other joint lines of business and of customer referrals, including future market and social relations akin to relational contracting, Gibbons and Roberts 2012). This yields a misconduct sanctioning vs. reputation-type interpretation.

Our goal is not to develop a general theory of either misconduct (e.g., Banerjee et al. 2012 for corruption) or reputation and moral hazard (e.g., Board and Meyer-ter-Vehn 2013). We rather provide a parsimonious model of moral hazard under revelation that embeds misconduct and sanctioning to deliver highly stylized predictions which guide the interpretation of our results.

A.1 Model: Misconduct, Punishment, and Reputation

A.1.1 Environment

We assume a continuum of local markets, defined by the pair (i, j), where i denotes a randomly selected vendor and j denotes potential customer(s). This is akin to our experiment's design, whereby we construct a local market using a randomly selected vendor and nearby households as customers per locality to maximize statistical power. In each locality, other vendors and customers have no designated role; our model will inherit the same design. We present a simple model of moral hazard under revelation with reputational effects and direct punishment.

The vendor chooses an action $s \in \{0, 1\}$, where $s = 0$ refers to a dishonest action (does overcharge market transaction) and $s = 1$ refers to an honest action (does not overcharge market transaction). Customers imperfectly observe the vendor's action, but learn about the transaction through public signals σ, giving rise to a moral hazard problem (Board and Meyer-ter-Vehn 2013). Denote by π the percentage of honest transactions (that is, the probability that the vendor will be honest), so $\Pr(s = 1) = \pi$. We allow customers to hold imperfect belief about the probability that the vendor will be honest, which we denote by $\hat{\pi}$. $\hat{\pi}$ is assumed to be common knowledge to avoid instances of higher-order beliefs.

The vendor receives revenue in two ways: reputation (from honest behavior) and "uncertain" direct benefits (from dishonest behavior). First, given public information σ, consumers' willingness to pay is $\mathbb{E}[\hat{s} = 1|\sigma]$; this equals the vendor's reputational payoff given the signal. We call this reputational payoff as the vendor cares about $\mathbb{E}[\hat{s} = 1|\sigma]$ that customers compute (i.e., posterior that the vendor is honest) and assigns 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. The vendor's reputational revenue is proportional to the market size (denoted by $\eta > 0$) and his/her belief that customers perceive his/her actions as honest. Second, if the vendor chooses $s = 0$ (a dishonest action), s/he receives an additional benefit $Y > 0$ corresponding to the overcharged transaction amount. However, with probability q, consumers can directly punish the vendor by reporting the dishonest behavior; the vendor gets $Y^r \mathbf{I}_{s=0} < Y \mathbf{I}_{s=0}$ if reported. Given the vendor's action s and market consumers' belief about this action $\hat{\pi}$, the vendor's profits $\mathbf{\Pi}(s)$ equal[16]

$$[qY^r + (1-q)Y]\,\mathbf{I}_{s=0} + \hat{\pi}\mathbb{E}[\hat{s} = 1|\sigma]\eta + (1-\hat{\pi})(1 - \mathbb{E}[\hat{s} = 1|\sigma])\eta$$

A.1.2 Mapping Model to Experiment

Before analyzing the framework, it is useful to discuss how our model and analysis map to our experimental setup. Market vendor(s) decide whether to commit misconduct ($s = 0$) or not ($s = 1$). Consumers (uniformed vs informed) learn about the transactional action through public signals σ. Based on their inference about a vendor's action given the available signal, a customer either assigns a reputational payoff ($\mathbb{E}[\hat{s} = 1|\sigma]$) to the vendor (via either PT or MR information programs) or reports the vendor's dishonest behavior as a direct punishment (via MR information program). If customers perceive (via $\hat{\pi}$) that the vendor is honest, then the vendor receives higher revenue (i.e., through repeated or large transactions and not being reported) and vice versa.

Our goal is to compare market information sets about misconduct: one "without" information and another "with" information assignment about misconduct. For the information assignment, we vary the information sets: one with technology to detect and reward misconduct behavior (reputation effects, where $\sigma = s$), another with technology to directly report

[16] $[qY^r + (1-q)Y]$ is the vendor's opportunity cost of being honest. Our simple sanctioning and reputation formulation provides a moral hazard analog of the labor supply and stigma (adverse selection) model of Bursztyn, González, and Yanagizawa-Drott (2020).

and punish misconduct behavior (reputation and punishment effects), and one with both. We model assignment of the anti-misconduct market information as either a shift in the distribution of $\hat{\pi}$ or $\mathbb{E}[\hat{s} = 1|\sigma]$. As we show (and as implied by the model), the information assignment (i) increases customers beliefs about the percentage of honest transactions $\hat{\pi}$; (ii) cause customers to update their beliefs about honest vendor behavior (thus to assign $\mathbb{E}[\hat{s} = 1|\sigma]$); and (iii) cause vendors themselves to update their beliefs about how informed consumers are and the likelihood of direct punishment. Together, these increase honest market vendor actions ($s = 1$) and improve market outcomes by increasing consumer demand for services and vendor sales revenue.

A.1.3 Analysis

In the game, we are interested in Perfect Bayesian Equilibria. Denote $\hat{\pi}^* = \frac{qY^r+(1-q)Y}{2\eta} + 1/2$ (assume $\hat{\pi}^* < 1$) (We provide detail foundations for η below).

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

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

This PBE is supported by the following beliefs:

- $\Pr(\hat{s} = 1) = \hat{\pi}$
- $\Pr(\hat{s} = 1|\sigma = s = 1, \hat{\pi} \geq \hat{\pi}^*) = 1$ and $\Pr(\hat{s} = 1|\sigma = s = 0, \hat{\pi} \geq \hat{\pi}^*) = 0$
- $\Pr(\hat{s} = 1|\sigma = s = 1, \hat{\pi} < \hat{\pi}^*) = \underbrace{x \in (0, 1)}$ and $\Pr(\hat{s} = 1|\sigma = s = 0, \hat{\pi} < \hat{\pi}^*) = \hat{\pi}$

Proof. See Proofs. ■

In our experiment, when we provide symmetric two-sided information about the official prices of transactions, consumers' signal σ is the same as the s action chosen by the vendor ($s = \sigma$). There is revelation of the imperfectly observed vendor's actions and beliefs are updated to the posterior $\Pr[\hat{s} = 1|\sigma = s = 1] = 1$ and $\Pr[\hat{s} = 1|\sigma = s = 0] = 0$. The maximal extent of reputation gain is given by the difference: $\triangle\mathbb{E}[\hat{s} = 1|\sigma] = \mathbb{E}[\hat{s} = 1|\sigma = s = 1] - \mathbb{E}[\hat{s} = 0|\sigma = s = 0]$ which depends on the available signal about the vendor's action σ 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 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 customers visits to the vendor, η, making equilibrium honest behavior $s = 1$ more likely. Informed consumers thus exert a positive externality on uninformed consumers by driving up honest vendor behavior.* (iii) *Increasing either $\triangle\mathbb{E}[\hat{s} = 1|\sigma]$ (PT or MR information programs) or q (MR information program) increases the number of $s = 1$.*

Proof. See Proofs. ■

A.2 Proof of Proposition 1

Proof. $s = 1$ IFF

$$\mathbf{\Pi}(s = 1) > \mathbf{\Pi}(s = 0)$$

$$-\frac{0}{\eta} + \hat{\pi}\mathbb{E}[\hat{s} = 1|\sigma, s = 1] + (1 - \hat{\pi})(1 - \mathbb{E}[\hat{s} = 1|\sigma, s = 1])$$
$$>$$
$$\frac{qY^s + (1-q)Y}{\eta} + \hat{\pi}\mathbb{E}[\hat{s} = 1|\sigma, s = 0] + (1 - \hat{\pi})(1 - \mathbb{E}[\hat{s} = 1|\sigma, s = 0])$$

$\mathbb{E}[\hat{s} = 1|\sigma, s] = \Pr[\hat{s} = 1|\sigma, s]$, so we write:

$$-\frac{0}{\eta} + \hat{\pi}\underbrace{\Pr[\hat{s} = 1|\sigma, s = 1]}_{\mu(1,1)} + (1 - \hat{\pi})(1 - \Pr[\hat{s} = 1|\sigma, s = 1])$$
$$>$$
$$\frac{qY^s + (1-q)Y}{\eta} + \hat{\pi}\underbrace{\Pr[\hat{s} = 1|\sigma, s = 0]}_{\mu(1,0)} + (1 - \hat{\pi})(1 - \Pr[\hat{s} = 1|\sigma, s = 0])$$

We get

$$2\hat{\pi}\mu(1,1) - 2\hat{\pi}\mu(1,0) - \mu(1,1) + \mu(1,0) > \frac{qY^s + (1-q)Y}{\eta}$$

$$\hat{\pi} > \frac{qY^s + (1-q)Y}{2\eta\triangle\mu} + 1/2$$

where $\triangle\mu = \mu(1,1) - \mu(1,0) = \triangle\mathbb{E}[\hat{s} = 1|\sigma]$. In this PBE:

If $\hat{\pi} > \hat{\pi}^*$, then $\mu(1,1) = \Pr(\hat{s} = 1|\sigma, s = 1) = 1$ and $\mu(1,0) = \Pr(\hat{s} = 1|\sigma, s = 0) = 0$. Since $\hat{\pi}$ is common knowledge, consumers calculate that if $\hat{\pi} > \hat{\pi}^*$, then $\triangle\mu = 1$ which assigns the maximum reputational revenue. Thus, $\triangle\mu = 1$, implying $\hat{\pi} > \frac{qY^s+(1-q)Y}{2\eta(1-0)} + 1/2 \geq \hat{\pi}^*$. If $\hat{\pi} < \hat{\pi}^*$, then $\mu(1,1) = \Pr(\hat{s} = 1|\sigma, s = 1) = x \in (0,1)$ (it can be anything), $\mu(1,0) = \Pr(\hat{s} = 1|\sigma, s = 0) = \hat{\pi}$, $\Delta\mu < 1$ and

$$\hat{\pi} < \hat{\pi}^* = \frac{qY^s + (1-q)Y}{2\eta(1-0)} + 1/2$$

The vendor does not find it worthwhile to choose an honest action $s = 1$ to seek for any reputation; not even the maximum reputation gain $\Delta\mu = (1-0) = 1$ makes it worthwhile to choose an honest action $s = 1$. The opportunity cost of being honest $[qY^s + (1-q)Y]$ is too high. In our experiment, by providing symmetric two-sided information about official and mandated prices of transactions, consumers' signal σ is the same as the s action chosen by the vendor ($s = \sigma$). There is revelation of the imperfectly observed vendor's actions and beliefs are updated to the posterior $\Pr[\hat{s} = 1|\sigma = s = 1] = 1$ and $\Pr[\hat{s} = 1|\sigma = s = 0] = 0$

. ■

A.3 Proof of Proposition 2

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

Foundations: Computing η: 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: $v(\Pr[\hat{s} = 1|\sigma, s]) = \Pr[\hat{s} = 1|\sigma, s]v_G + (1 - \Pr[\hat{s} = 1|\sigma, s])v_B$, with a reduced form demand function: $D_i(\Pr[\hat{s} = 1|\sigma, s] = 1) = v(\Pr[\hat{s} = 1|\sigma, s] = 1) = \theta v_G$ for informed customers versus $D_u(\Pr[\hat{s} = 1|\sigma, s]) = v(\Pr[\hat{s} = 1|\sigma, s]) = (1-\theta)v(\Pr[\hat{s} = 1|\sigma, s])$ for uninformed customers. Thus, the aggregate market demand for honest transactions is

$$D_{s=1}(\Pr[\hat{s} = 1|\sigma, s]) = \underbrace{\theta v_G}_{D_i} + \underbrace{(1-\theta)v(\Pr[\hat{s} = 1|\sigma, s])}_{D_u}$$

Similarly, the aggregate demand is $D_{s=0}(\Pr[\hat{s} = 1|\sigma, s]) = \theta v_B + (1-\theta)v(\Pr[\hat{s} = 1|\sigma, s])$ for dishonest transactions.

Effects: Letting η equal the aggregate demand D_s, and observing that $\frac{\partial D_{s=1}}{\partial \theta} = v_G - v(\Pr[\hat{s} =$

$1|\sigma, s]) = v_G - \Pr[\hat{s} = 1|\sigma, s]v_G - (1 - \Pr[\hat{s} = 1|\sigma, s])v_B \geq 0|_{\Pr[\hat{s}=1|\sigma,s]=1}$ in equilibrium. For dishonest transactions, $\frac{\partial D_{s=0}}{\partial \theta} = v_B - v(\Pr[\hat{s} = 1|\sigma, s]) \leq 0|_{\Pr[\hat{s}=1|\sigma,s]=1}$. We thus have the following result: For (ii), $\eta(\theta)$ *is weakly-increasing in* θ*. Since* $\hat{\pi}^*$ *is decreasing in* η*, noting that* $\lim_{\eta \to +\infty} \hat{\pi}^* = 0$*, it follows that* $\Pr(s = 1)$ *is more likely.*

To prove (iii), it suffice to show that $\frac{\partial \hat{\pi}}{\partial q}|_{\hat{\pi}=\hat{\pi}^{**}} < 0$ and $\frac{\partial \hat{\pi}}{\partial \triangle\mu}|_{\hat{\pi}=\hat{\pi}^{**}} < 0$ where $\hat{\pi}^{**} = \frac{qY^s+(1-q)Y+0}{2\eta\triangle\mu} + 1/2$ since both make $\Pr(s = 1)$ more likely. We have that $\frac{\partial \hat{\pi}}{\partial q}|_{\hat{\pi}=\hat{\pi}^{**}} = \frac{Y^s - Y}{2\eta\triangle\mu} < 0$ because $Y^s < Y$. Similarly, $\frac{\partial \hat{\pi}}{\partial \triangle\mu}|_{\hat{\pi}=\hat{\pi}^{**}} = -\frac{2\eta(qY^s+(1-q)Y+0)}{(2\eta\triangle\mu)^2} < 0$. ∎

B Balance, Attrition, and Descriptive Statistics

Table B.1: **Balance test I: representativeness of select-sample with market population**

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***	8.96*e*-04
	(0.054)	(0.076)
Age (years)	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		
Head of household reads English	4.104***	0.102
	(0.163)	(0.223)
Outer wall uses cement	3.909***	-0.306
	(0.222)	(0.342)
Toilet facility	4.617***	-0.349
	(0.140)	(0.268)
Number of 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: No. of Customers + Sales Revenue		
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 B.2: **Balance test i: representativeness of select-sample with market population**

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 (years)	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		
Head of household reads English	3.428***	-0.124
	(0.114)	(0.152)
Outer wall uses cement	3.664***	-0.272
	(0.196)	(0.195)
Toilet facility	4.372***	-0.584
	(0.137)	(0.182)
Number of 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 Beliefs: Vendor Misconduct		
Attempted fraud experience (any)	0.611***	-0.041
	(0.040)	(0.039)
Ever overcharged/unauthorized account use	0.292***	0.013
	(0.024)	(0.028)
Market: Features + Transactions / Demand		
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 customer 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 B.3: **Balance test II: pre-intervention treatment-control differences**

SUPPLY SIDE: VENDORS

	Constant	**PT**	**MR**	**Combined: PT+MR**
Demographic Characteristics				
Female	0.560***	-0.196	-0.263*	-0.078
	(0.179)	(0.156)	(0.142)	(0.159)
Married	0.234*	-0.0517	-0.207	-0.144
	(0.129)	(0.145)	(0.135)	(0.131)
Akan ethnic	0.340***	0.175	-0.137	0.158
	(0.158)	(0.137)	(0.147)	(0.130)
Age (years)	25.50***	-0.664	2.162	-1.535
	(2.166)	(2.592)	(2.378)	(2.132)
Education (any)	0.752***	-0.018	0.057	-0.035
	(0.225)	(0.171)	(0.158)	(0.151)
Self-employment	0.462	0.055	0.027	-0.125
	(0.186)	(0.159)	(0.156)	(0.142)
M-Money training	0.243	0.266	0.258*	0.151
	(0.238)	(0.162)	(0.154)	(0.135)
Poverty Indicators				
Head of household reads English	4.613***	-0.047	-0.120	0.271
	(0.363)	(0.483)	(0.458)	(0.429)
Outer wall uses cement	2.965***	0.154	-0.205	-0.529
	(0.996)	(0.753)	(0.681)	(0.710)
Toilet facility	4.684***	0.522	-0.451	-0.461
	(0.426)	(0.570)	(0.651)	(0.542)
Number of working mobile phones	9.039***	-0.108	0.381	-0.350
	(0.434)	(0.493)	(0.487)	(0.445)
Own working bicycle/motor bicycle/car	0.785	-0.076	0.268	0.318
	(0.821)	(0.958)	(0.963)	(0.956)
Poverty rate (Schreiner 2015), %	0.261	4.339	0.575	3.793
	(3.489)	(5.942)	(4.287)	(4.039)
Market: No. of Customers + Sales Revenues + Misconduct				
M-Money: Total volume [GHS] (daily)	1783	373.4	554.3	640.0
	(1206)	(851.0)	(975.8)	(1487)
Non-M-Money: Number customers (daily)	26.65***	-2.246	-7.728	9.261
	(8.283)	(8.325)	(9.145)	(13.09)
Non-M-Money: Total volume [GHS] (daily)	202.0***	-26.12	9.998	26.65
	(70.50)	(57.60)	(60.18)	(69.88)
Misconduct amount [GHS] -- Audit Measure	-0.031	0.261	-0.062	-0.011
	(0.141)	(0.238)	(0.197)	(0.191)
Joint F-test (linear), p-value			0.711	
Chi-squared test (probit), p-value			0.534	

Note: Observations are at the vendor level. Each row is a separate regression and controls for randomization strata/district dummies. The F and Chi-squared tests are conducted using the pooled 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. Baseline outcomes (sales revenue, misconduct) and characteristics are balanced across treatment arms.

Table B.4: **Balance test II: pre-intervention treatment-control differences**

DEMAND SIDE: CUSTOMERS

	Constant	**PT**	**MR**	**Combined: PT+MR**
Demographic Characteristics				
Female	0.713***	0.001	-0.001	-0.030
	(0.059)	(0.056)	(0.064)	(0.058)
Married	0.489***	0.028	-0.001	0.074
	(0.058)	(0.047)	(0.045)	(0.052)
Akan ethnic	0.615***	0.058	0.057	0.071
	(0.092)	(0.083)	(0.090)	(0.085)
Age (years)	42.99***	1.541	0.253	0.644
	(1.914)	(1.670)	(1.843)	(1.677)
Education (any)	0.888***	0.028	-0.027	0.019
	(0.034)	(0.027)	(0.039)	(0.031)
Self-employment	0.708***	0.018	0.057	0.033
	(0.044)	(0.049)	(0.062)	(0.054)
M-Money registered	0.750***	-0.011	-0.005	0.002
	(0.045)	(0.031)	(0.030)	(0.032)
Poverty Indicators				
Head of household reads English	2.921***	0.055	-0.231	0.288
	(0.256)	(0.236)	(0.251)	(0.193)
Outer wall uses cement	2.346***	-0.146	0.262	0.270
	(0.654)	(0.442)	(0.447)	(0.452)
Toilet facility	4.487***	-0.318	-0.308	-0.575*
	(0.201)	(0.308)	(0.373)	(0.293)
Number of working mobile phones	6.959***	-0.421	-0.055	0.073
	(0.362)	(0.289)	(0.289)	(0.299)
Own working bicycle/motor bicycle/car	0.533	0.145	0.393	0.466
	(0.321)	(0.315)	(0.340)	(0.388)
Poverty rate (Schreiner 2015), %	12.696***	1.973	0.873	-0.127
	(1.739)	(1.960)	(2.010)	(1.739)
Subjective Beliefs: Vendor Misconduct				
Attempted fraud experience (any)	0.454***	-0.020	0.008	-0.030
	(0.068)	(0.066)	(0.064)	(0.065)
Ever overcharged/unauthorized account use	0.222***	-0.080	-0.054	-0.025
	(0.062)	(0.049)	(0.049)	(0.052)
Market: Features + Transactions / Demand				
Distance to closest formal bank (meters)	369.2*	45.59	195.2	415.0**
	(188.2)	(123.1)	(233.6)	(189.6)
Distance to closest M-Money (meters)	16.97	26.21	-7.900	-0.960
	(11.30)	(16.55)	(15.50)	(14.01)
M-Money: Total use volume [GHS] (weekly)	184.0***	-29.29	-5.371	43.73
	(63.79)	(39.50)	(39.15)	(54.56)
Non-M-Money: Number use (weekly)	2.400*	-0.279	1.079	0.611
	(1.237)	(0.728)	(1.772)	(1.160)
Non-M-Money: Total use volume [GHS] (weekly)	24.28	31.22	18.69	18.75
	(30.26)	(28.40)	(17.46)	(21.80)
Borrowing + Savings				
Likelihood to borrow via M-Money (1-5 scale)	1.251***	-0.020	0.054	0.094
	(0.174)	(0.126)	(0.157)	(0.167)
Likelihood to save via M-Money (1-5 scale)	1.779***	-0.122	-0.055	-0.017
	(0.149)	(0.159)	(0.157)	(0.175)
Joint F-test (linear), *p*-value			0.850	
Chi-squared test (probit), *p*-value			0.846	

Note: Observations are at the customer level. Each row is a separate regression and controls for randomization strata/district dummies. The F and Chi-squared tests are conducted using the pooled 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. Baseline outcomes (beliefs, demand, poverty) and characteristics are balanced across treatment arms.

Table B.5: **ATTRITION**

	PT	**MR**	**Combined: 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 surveys*)						
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 summary statistics for the subsample that was successfully reached for follow-up and for the subsample that was not successfully reached in the endline survey-based exercises. Shown for both sides of the market (vendors versus customers). We fail to reject the null that attrition is non-differential. Attrition for endline audit exercises is 0.8%: 129 out of the 130 randomly selected vendors were reached. There was only one unreachable vendor in the combined PT + MR program.

Table B.6: **ATTRITION – TEST FOR SIGNIFICANCE BY TREATMENT PROGRAM**

CUSTOMERS DV: Customer dropped out or unreachable at endline indicator		
	(1)	(2)
Any treatment	-0.055	
	(0.038)	
Transparency alone		-0.072
		(0.044)
Monitoring alone		-0.032
		(0.047)
Combined		-0.060
		(0.043)
Constant	0.227***	0.227***
	(0.035)	(0.035)
Observations	990	990
Mean of dependent variable in control	0.227	0.227

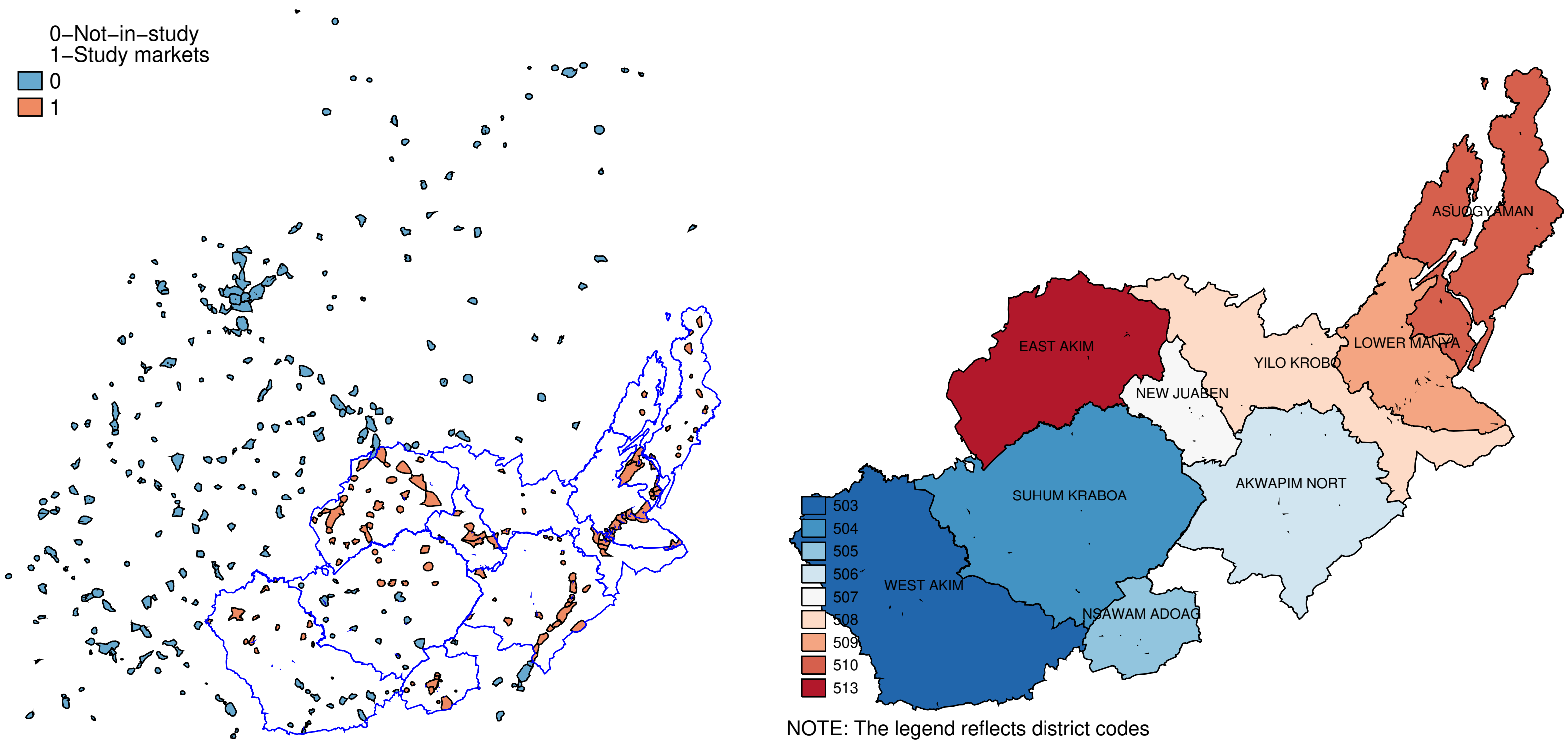
Note: Table shows differences in attrition rate (0-1 indicator for unreachable customers at endline). Observations are at customer level. Cluster-robust standard errors at the locality level are in parentheses *** p<0.01, ** p<0.05, * p<0.10. Attrition is non-differential (i) between pooled treatment arm and control arm and (ii) between the separate treatment arms and control arm at conventional significance levels.

Table B.7: **ATTRITION – TEST FOR SIGNIFICANCE BY TREATMENT PROGRAM**

VENDORS DV: Vendor dropped out or unreachable at endline indicator		
	(1)	(2)
Any treatment	-0.055	
	(0.082)	
Transparency alone		-0.057
		(0.100)
Monitoring alone		-0.093
		(0.095)
Combined		-0.018
		(0.101)
Constant	0.218***	0.218***
	(0.073)	(0.074)
Observations	130	130
Mean of dependent variable in control	0.218	0.218

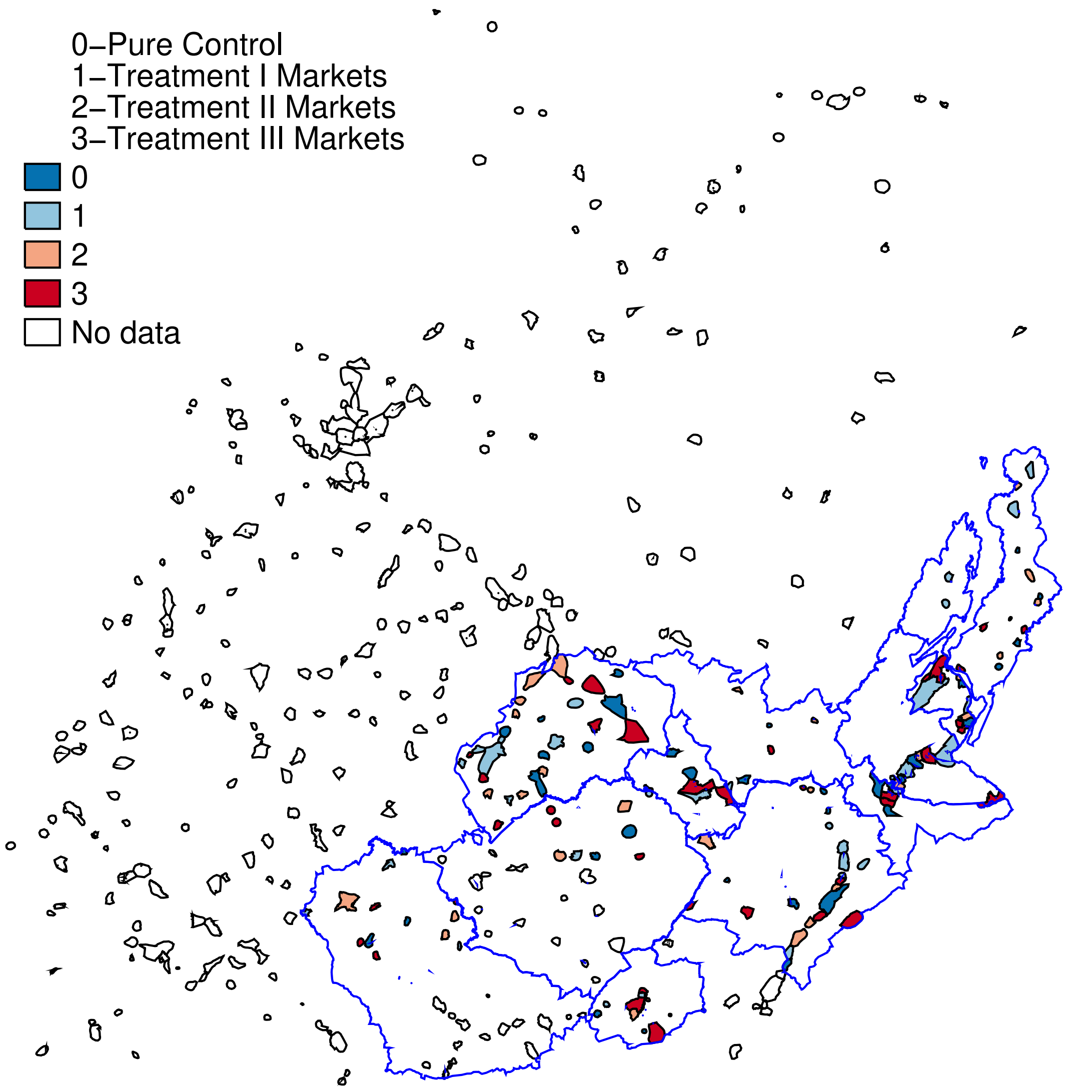
Note: Table shows differences in attrition rate (0-1 indicator for unreachable vendors at endline). Observations are at vendor level. Heteroskedasticity-robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.10. Attrition is non-differential (i) between pooled treatment arm and control arm and (ii) between the separate treatment arms and control arm at conventional significance levels.

Figure B.1: **Map for market census – spatial distribution of local markets**



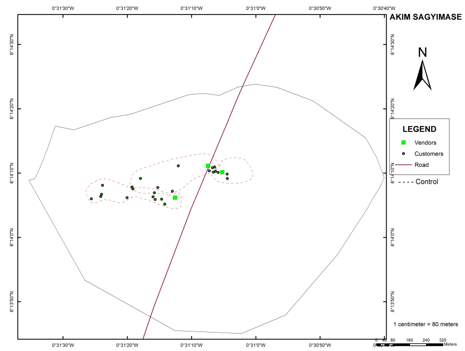
Note: Figure (left) shows the spatial distribution of localities in the study area (the Eastern belt of Ghana). The small polygons reflect localities. Selected localities (n=130) for the baseline market census are located in 9 administrative districts, namely: West Akim, Nsawam Adoagyiri, Suhum Kraboa, East Akim, New Juaben, Akwiapim North, Yilo Krobo, Lower Manya Krobo, and Asuogyaman. The district boundaries are displayed and projected in the figure (right). To build the market censuses, we initially restrict attention to localities that have a total population between 1000-20,000 people (mean=3900 people and median=2300 people) to maximize the chance of having a M-Money vendor present in the locality.

Figure B.2: **Map for treatment assignments at the local market level**

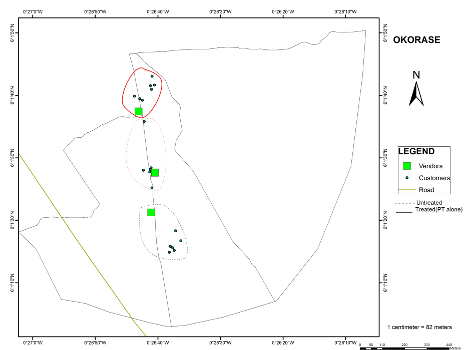


Note: Figure shows the spatial distribution of localities by treatment status. The polygons reflect the select localities (n=130). We stratify treatment programs based on the 9 administrative districts, namely: West Akim, Nsawam Adoagyiri, Suhum Kraboa, East Akim, New Juaben, Akwiapim North, Yilo Krobo, Lower Manya Krobo, and Asuogyaman (district boundaries are displayed). We are able to identify distinct markets, which limits potential cross-market spillovers: (i) As displayed, most localities are spatially distinct and (ii) Consumers report not switching to use different vendors other than the nearby, local vendors. We collected transaction networks data pre and post the interventions: both treated and control consumers were shown the market-level roster of vendors and then asked to indicate which vendor they last transacted with. The probability of repeat visits is not differential between the treated and control consumers (p-values > 0.20 in a regression of the probability of repeat visits against the treatment dummies with errors clustered at the locality level). This is inconsistent with cross-market spillovers.

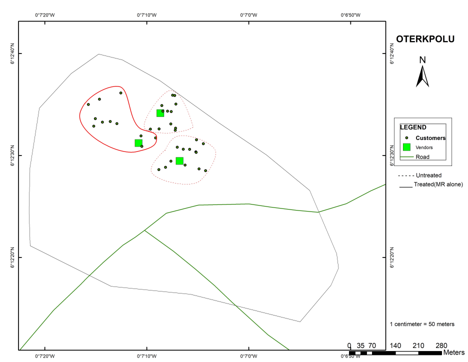
Figure B.3: **MAPS FOR SELECT MARKETS SHOWING PARTICIPANTS LOCATIONS**



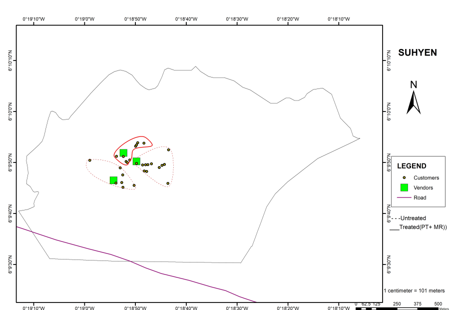
(a) MARKET: AKIM SAGYIMASE



(b) MARKET: OKORASE



(c) MARKET: OTERKPOLU



(d) MARKET: SUHYEN

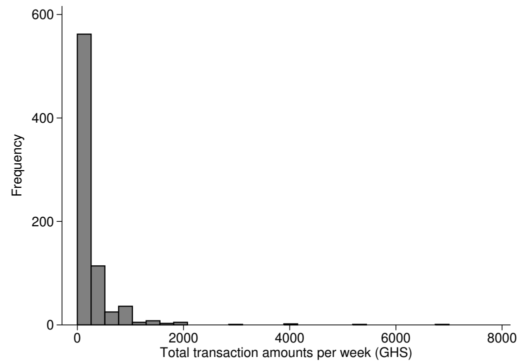
Note: Figure shows the locations of treated and untreated vendors and surveyed households / customers across 4 select markets: Akim Sagyimase (Control locality), Okorase (Price Transparency PT-alone locality), Oterkpolu (Monitoring MR-alone locality), and Suhyen (Combined PT+MR locality). Each of these markets has 3 vendors. In treated markets, only one random vendor and nearby households are treated.

Table B.8: **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)	26.29	8.242	39.54	15.02
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
Head of household reads English	0.769	0.421	0.606	0.488
Outer wall uses 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: Number customers (daily)	42.93	45.13		
M-Money: Total volume [GHS] (daily)	2260	3775		
Non-M-Money: Number customers (daily)	32.79	47.06		
Non-M-Money: Total volume [GHS](daily)	155.1	164.5		
Distance to closest formal bank (meters)			338.5	751.3
Distance to closest post office (meters)			382.9	250.7
Distance to closest M-Money (meters)			61.28	94.92
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)			144.1	396.2
Non-M-Money: Number use (weekly)			2.272	14.76
Non-M-Money: Total use volume [GHS] (weekly)			44.70	505.1
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 overcharged			0.191	0.403
Ever overcharged + unauthorized account use			0.293	0.455
Number of observations	333		1,921	

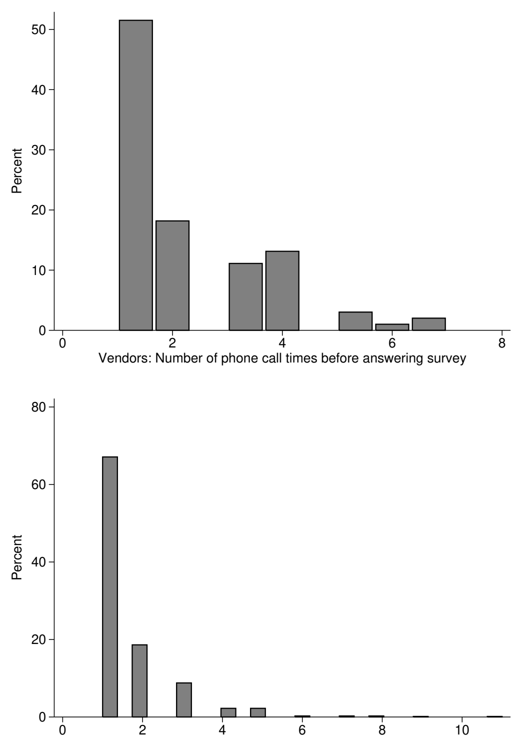
Note: Table reports summary statistics of relevant variables from our market census separately for both sides of the market: vendors *versus* customers. This includes information about demographics, poverty indicators, and market outcomes, respectively. Customers' borrowing and savings behavior and their subjective assessment of market misconduct on M-Money are also shown. The census covers 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 B.4: **Distribution (histogram) of total transactions at endline**



Observations are at the customer level.

Figure B.5: **Phone calls and reachability of subjects**



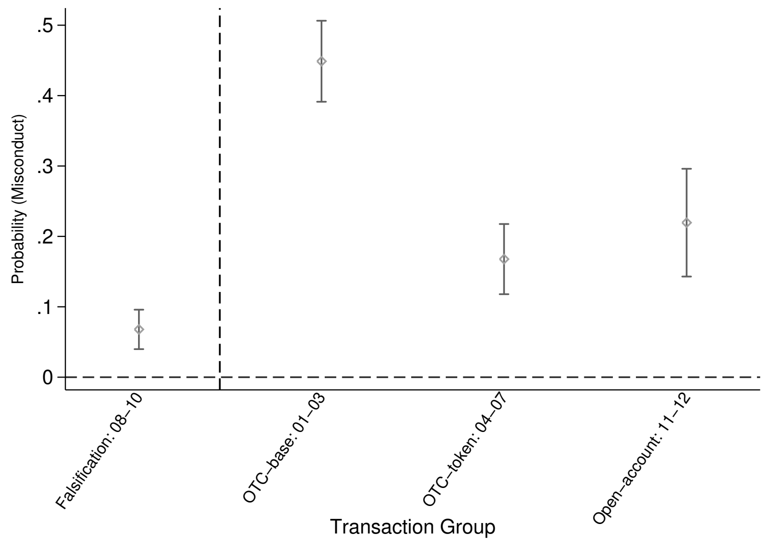
Observations are at the subject (vendor, customer, respectively) level.

Table B.9: **Misconduct at baseline: descriptive statistics based on transactional audit exercise, details**

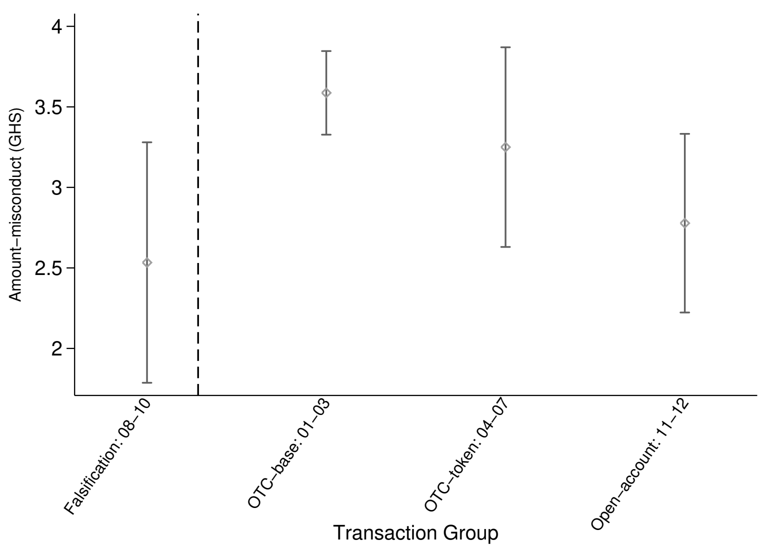
# Transaction type (description)	Outcome variable	Mean	SD	Transaction group	Mean	SD
01 Cash-in GHS50 - to others' wallet	**1**[Misconduct=Yes]	0.35	0.480		0.44	0.498
	Overcharged [GHS]	4.65	1.093		3.58	1.498
02 Cash-in GHS160 - to others' wallet	**1**[Misconduct=Yes]	0.52	0.502	$\{= \quad OTC-base$		
	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 successful transactions		663			663	

Note: Table reports the specific transactions used for the actual transactional exercises and shows the descriptive statistics of misconduct (n=663). These misconduct outcomes are based on the transactional exercises. Transactions are categorized into four groups: 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., leave 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 amount overcharged 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 three varying transaction sizes, is approximately GHS4.03.

Figure B.6: **Misconduct at baseline: descriptive statistics based on transactional audit exercise, graphical**



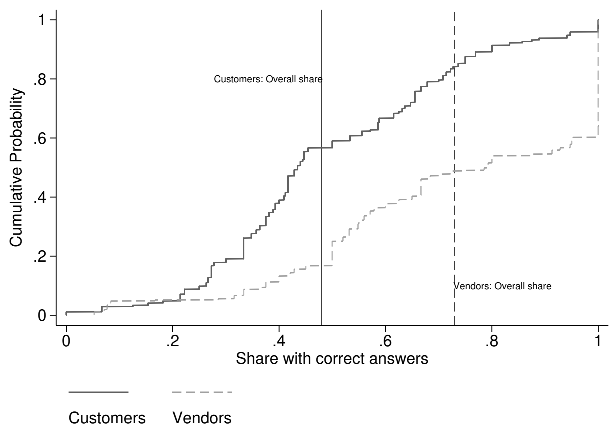
(a) **MISCONDUCT INCIDENCE × TRANSACTION GROUP**



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

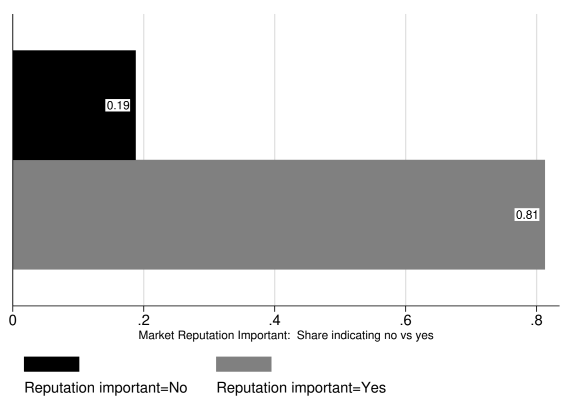
Note: Figures display the distribution of misconduct (n=663), measured as either the probability of the vendor committing a misconduct "incidence" (Figure (a)) or the amount overcharged as result of misconduct "severity" (Figure (b)) using actual transactional exercises at baseline. Transactions are categorized into four groups: 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 Table B.9. 90% confidence intervals (CI) are displayed around the estimates. As expected, misconduct is much higher in OTC-type transactions (i.e., little to no automation/verification required from the customer) compared to the Falsification group (automation and active verification required from the customer).

Figure B.7: **Asymmetric information about transactional prices**



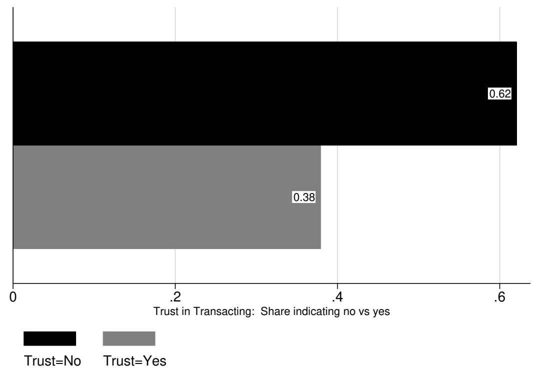
Note: Figure plots the distributions (CDFs) of the share of subjects with accurate answers for charges on randomly selected popular transactions (GHS200; GHS1200) derived with reference to their official or mandated rates (2GHS; 10GHS, respectively). A subject (customer, vendor) is correct if his/her answer matches the mandated rate. Observations are at the subject level. In each local market, we compute the share of subjects who answered correctly. Shown separately for customers and vendors. Trimmed to exclude unrealistic zero vendor knowledge/ correctness at the local market level. From a Kolmogorov–Smirnov test for the equality of distributions, p-value < 0.01.

Figure B.8: **Importance of reputation to vendors**



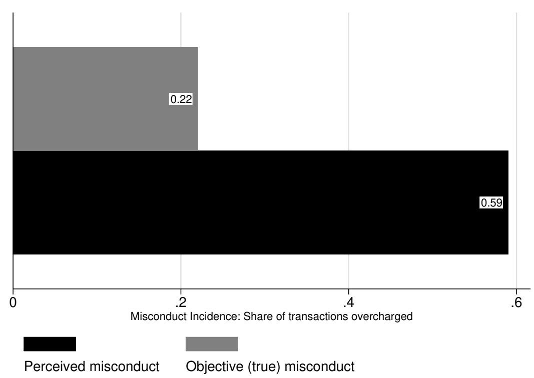
Note: Figure plots the share of vendors who value good market reputation through their money market transactions. Subjects (vendors) are asked to indicate how important it is to show a high degree of good market image and responsibility to potential customers when carrying out M-Money transactions on 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 B.9: **Consumer trust in performing money transactions at vendor points**



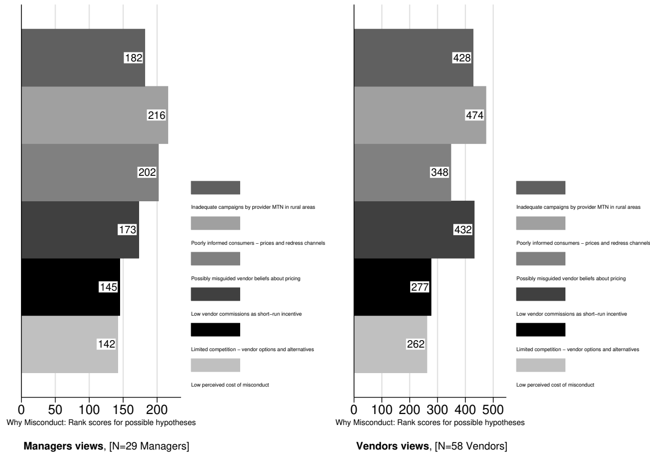
Note: Figure plots the share of customers, at baseline, who trust or do not trust the money transactions they make at vendor banking points. Subjects (customers) are asked to indicate their level of trust for carrying out M-Money transactions at vendor points from a scale of 1 (low) to 5 (high). To ease 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 B.10: **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 transactions not overcharged.

Figure B.11: **High misconduct: follow-up surveys with managers and vendors**



Note: Figure shows the overall rankings from managers of the service provider and vendors in control markets for 9 hypotheses in order of the most plausible reason for "why vendor misconduct is prevalent in low-income areas". We recoded the rankings to get a rank of 1 achieve the highest score. For each hypothesis, we calculate the total ranking (which equals the sum of rank scores) across all respondents. We then plot the total score for each hypothesis, which is captured by the height of the horizontal bars and displayed in each bar to ease the exposition. Overall rankings are invariant to different aggregation approaches: median score vs mean score.

C Robustness Checks and Further Results

C.1 Robustness Checks: Inference, Multiple Testing, Attrition, LASSO Estimation

Seller Misconduct and Consumers' Beliefs (1)

Table C.1: **Prices: effect of information sets on vendor misconduct**

	Misconduct indicator	Misconduct amount (GHS)
	(1)	(2)
Any treatment	-0.211	-0.550
	(0.086)	(0.255)
CI: Clustered S.E.	[-0.382, -0.039]	[-1.059, -0.042]
Inference Robustness		
p-value: Wild Bootstrap	0.017	0.016
p-value: Permutation Test	0.001	0.009
p-value: R-W MHT Corr (2005)	0.042	0.054
Attrition Robustness		
Lee (2009) Attrition Bounds	<-0.174, -0.164>	<-0.484, -0.435>
Imbens and Manski (2004) CS	[-0.235, -0.081]	[-0.675, -0.013]
Observations	335	335
Mean of dep var in control	0.294	0.778

Note: Observations are at the select vendor x transaction type x transaction date level. Dependent variables are audit-based measures. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at vendor level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for vendor-related outcomes family (misconduct 0/1; misconduct amount) and jointly includes both the pooled and separate treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) and confidence sets (CS) in brackets. Results are similar to exclusion of controls and (ii) to post-double-selection LASSO estimates clustered at the vendor level.

Table C.2: **Prices: effect of information sets on vendor misconduct**

	Misconduct indicator	Misconduct amount (GHS)
	(1)	(2)
Transparency alone	-0.184	-0.439
	(0.094)	(0.276)
CI: Clustered S.E.	[-0.372, -0.003]	[-0.988, 0.110]
p-value: Wild Bootstrap	0.045	0.093
p-value: Permutation Test	0.004	0.033
p-value: R-W MHT Corr (2005)	0.105	0.150
Lee (2009) Attrition Bounds	<-0.084, -0.069>	<-0.295, -0.213>
Imbens and Manski (2004) CS	[-0.129, 0.017]	[-0.416, 0.218]
Monitoring alone	-0.217	-0.574
	(0.093)	(0.275)
CI: Clustered S.E.	[-0.403, -0.031]	[-1.122, -0.027]
p-value: Wild Bootstrap	0.012	0.022
p-value: Permutation Test	0.000	0.000
p-value: R-W MHT Corr (2005)	0.038	0.052
Lee (2009) Attrition Bounds	<-0.132, -0.064>	<-0.409, -0.131>
Imbens and Manski (2004) CS	[-0.214, -0.024]	[-0.513, 0.002]
Combined	-0.211	-0.554
	(0.089)	(0.279)
CI: Clustered S.E.	[-0.390, -0.033]	[-1.110, -0.001]
p-value: Wild Bootstrap	0.013	0.022
p-value: Permutation Test	0.000	0.000
p-value: R-W MHT Corr (2005)	0.045	0.063
Lee (2009) Attrition Bounds	<-0.029, 0.034>	<-0.070, 0.217>
Imbens and Manski (2004) CS	[-0.073, 0.117]	[-0.205, 0.549]
Observations	335	335
Mean of dep var in control	0.294	0.778

Note: Observations are at the select vendor x transaction type x transaction date level. Dependent variables are audit-based measures. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at vendor level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for vendor-related outcomes family (misconduct 0/1; misconduct amount) and jointly includes both the pooled and separate treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) and confidence sets (CS) in brackets. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates clustered at the vendor level.

Table C.3: **Consumers' beliefs about vendor honesty**

	Belief about vendor honesty indicator	
	(1)	(2)
Any treatment	0.070	
	(0.040)	
CI: Clustered S.E.	[-0.011, 0.145]	
p-value: Wild Bootstrap	0.129	
p-value: Permutation Test	0.117	
p-value: R-W MHT Corr ('05)	0.007	
Lee (2009) Attrition Bounds	<0.056, 0.128>	
Imbens and Manski (2004) CS	[-0.028, 0.208]	
Transparency alone		0.107
		(0.057)
CI: Clustered S.E.		[0.007, 0.221]
p-value: Wild Bootstrap		0.082
p-value: Permutation Test		0.056
p-value: R-W MHT Corr ('05)		0.035
Lee (2009) Attrition Bounds		<0.021, 0.067>
Imbens and Manski (2004) CS		[-0.052, 0.138]
Monitoring alone		-0.045
		(0.057)
CI: Clustered S.E.		[-0.158, 0.068]
p-value: Wild Bootstrap		0.462
p-value: Permutation Test		0.991
p-value: R-W MHT Corr ('05)		0.351
Lee (2009) Attrition Bounds		<-0.080, -0.059>
Imbens and Manski (2004) CS		[-0.154, 0.019]
Combined		0.126
		(0.054)
CI: Clustered S.E.		[0.018, 0.233]
p-value: Wild Bootstrap		0.038
p-value: Permutation Test		0.000
p-value: R-W MHT Corr ('05)		0.002
Lee (2009) Attrition Bounds		<0.076, 0.102>
Imbens and Manski (2004) CS		[0.001, 0.175]
Observations	810	810
Mean of dep var in control	0.314	0.314

Note: Observations are at the customer level. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Belief denotes customers' perception that they are not being overcharged at vendor points (or perception that they have not experienced seller misconduct) at endline. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at locality level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for customer-related outcomes family (belief about vendor honesty; beliefs update about vendor misconduct) and jointly includes both the pooled, separate and interactions with treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) and confidence sets (CS) in brackets. Results are similar (i) to exclusion of controls, (ii) to post-double-selection LASSO estimates clustered at the locality level, and (iii) to alternative beliefs measures: non-incentivized versus incentivized outcomes. Omitted to conserve space.

Table C.4: **Consumers' belief update about vendor misconduct**

	Belief about vendor misconduct indicator		Belief about vendor misconduct indicator
	(1)		(2)
Any treatment	-0.282	**Transparency alone**	-0.365
	(0.082)		(0.087)
CI: Clustered S.E.	[-0.445, -0.119]	CI: Clustered S.E.	[-0.537, -0.192]
x Objective misconduct	0.273	**x Objective misconduct**	0.349
	(0.106)		(0.122)
CI: Clustered S.E.	[0.062, 0.483]	CI: Clustered S.E.	[0.107, 0.592]
p-value: Wild Bootstrap	0.095	*p*-value: Wild Bootstrap	0.047
p-value: Permutation Test	0.000	*p*-value: Permutation Test	0.000
p-value: R-W MHT Corr ('05)	0.027	*p*-value: R-W MHT Corr ('05)	0.015
Lee (2009) Attrition Bounds	%s exactly equal	Lee (2009) Attrition Bounds	%s exactly equal
Imbens and Manski (2004) CS	%s exactly equal	Imbens and Manski (2004) CS	%s exactly equal
		Monitoring alone	-0.152
			(0.093)
		CI: Clustered S.E.	[-0.338, 0.033]
		x Objective misconduct	0.235
			(0.121)
		CI: Clustered S.E.	[-0.004, 0.475]
		p-value: Wild Bootstrap	0.128
		p-value: Permutation Test	0.000
		p-value: R-W MHT Corr ('05)	0.092
		Lee (2009) Attrition Bounds	%s exactly equal
		Imbens and Manski (2004) CS	%s exactly equal
		Combined	-0.354
			(0.078)
		CI: Clustered S.E.	[-0.510, -0.199]
		x Objective misconduct	0.284
			(0.109)
		CI: Clustered S.E.	[0.067, 0.501]
		p-value: Wild Bootstrap	0.068
		p-value: Permutation Test	0.000
		p-value: R-W MHT Corr ('05)	0.029
		Lee (2009) Attrition Bounds	%s exactly equal
		Imbens and Manski (2004) CS	%s exactly equal
Objective misconduct	-0.199	Objective Misconduct	-0.216
	(0.087)		(0.082)
	[-0.373, -0.0255]		[-0.380, -0.053]
Observations	810		810
Mean of dep var (control)	0.685		0.685

Note: Observations are at the customer level. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Belief denotes customers' perception that they are being overcharged at vendor points (or perception that they have experienced seller misconduct) at endline. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at locality level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for customer-related outcomes family (belief about vendor honesty; beliefs update about vendor misconduct) and jointly includes both the pooled, separate and interactions with treatment indicators (includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) and confidence sets (CS) in brackets. Results are similar (i) to exclusion of controls, (ii) to post-double-selection LASSO estimates clustered at locality level, and (iii) to alternative beliefs measures: non-incentivized versus incentivized outcomes. Omitted to conserve space.

Consumers' Use of M-Money and Savings (2)

Table C.5: **Quantities: effect of information sets on usage of services and savings**

	`asinh` (Total transactions per week)	Used M-Money (last month)	Saved (last month)	PCA Index (1, 2, 3)
	(1)	(2)	(3)	(4)
Any treatment	0.458	0.073	0.075	0.188
	(0.225)	(0.039)	(0.042)	(0.091)
CI: Clustered S.E.	[0.011, 0.905]	[0.004, 0.151]	[0.008, 0.158]	[0.008, 0.368]
Inference Robustness				
p-value: Wild Bootstrap	0.055	0.092	0.087	0.062
p-value: Permutation Test	0.032	0.028	0.065	0.018
p-value: R-W MHT Corr (2005)	0.041	0.041	0.022	0.034
Attrition Robustness				
Lee (2009) Attrition Bounds	<0.366, 0.935>	<0.366, 0.935>	<0.070, 0.125>	<0.181, 0.243>
Behaghel et al. (2015) Attrition Bounds	<0.489, 0.756>	<0.090, 0.122>	<0.085, 0.117>	<0.214, 0.309>
Observations	810	810	810	810
Mean of dependent variable (control)	4.096	0.734	0.622	-0.201

Note: Observations are at the customer level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Cluster-robust standard errors at the market (locality) level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at market level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for consumer-related outcomes family (transaction amount; 0-1 usage of services; 0-1 saved on M-Money) and jointly includes both the pooled and separate treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) in brackets. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates to post-double-selection LASSO estimates clustered at the locality level.

Table C.6: **Quantities: effect of information sets on usage of services and savings**

	`asinh` (Total transactions per week)	Used M-Money (last month)	Saved (last month)	PCA Index (1, 2, 3)
	(1)	(2)	(3)	(4)
Transparency alone	0.262	0.048	0.047	0.118
	(0.263)	(0.044)	(0.047)	(0.106)
p-value: Clustered S.E.	[-0.260, 0.784]	[-0.040, 0.137]	[-0.046, 0.141]	[-0.091, 0.329]
CI: Wild Bootstrap	0.351	0.308	0.352	0.292
p-value: Permutation Test	0.233	0.149	0.282	0.161
p-value: R-W MHT Corr (2005)	0.499	0.499	0.499	0.499
Lee (2009) Attrition Bounds	<-0.328, 0.274>	<-0.034, 0.012>	<-0.054, -0.007>	<-0.115, -0.070>
Behaghel et al. (2015) Attrition Bounds	<-0.534, 0.251>	<09.055, 0.049>	<-0.067, 0.036>	<-0.178, 0.112>
Monitoring alone	0.587	0.084	0.042	0.223
	(0.268)	(0.044)	(0.052)	(0.104)
p-value: Clustered S.E.	[0.056, 1.119]	[-0.004, 0.172]	[-0.061, 0.146]	[0.016, 0.430]
CI: Wild Bootstrap	0.040	0.073	0.439	0.042
p-value: Permutation Test	0.000	0.000	0.001	0.000
p-value: R-W MHT Corr (2005)	0.064	0.076	0.414	0.064
Lee (2009) Attrition Bounds	<0.274, 0.449>	<0.038, 0.060>	<-0.023, -0.002>	<0.140, 0.162>
Behaghel et al. (2015) Attrition Bounds	<0.051, 0.943>	<0.044, 0.155>	<-0.046, 0.065>	<0.061, 0.382>
Combined	0.540	0.087	0.131	0.226
	(0.255)	(0.043)	(0.048)	(0.102)
p-value: Clustered S.E.	[0.035, 1.046]	[0.001, 0.174]	[0.035, 0.227]	[0.024, 0.428]
CI: Wild Bootstrap	0.042	0.061	0.009	0.035
p-value: Permutation Test	0.000	0.000	0.000	0.000
p-value: R-W MHT Corr (2005)	0.032	0.032	0.009	0.022
Lee (2009) Attrition Bounds	<0.182, 0.388>	<0.037, 0.063>	<0.107, 0.132>	<0.091, 0.112>
Behaghel et al. (2015) Attrition Bounds	<-0.004, 0.739>	<0.027, 0.125>	<0.095, 0.193>	<0.041, 0.313>
Observations	810	810	810	810
Mean of dependent variable (control)	4.096	0.734	0.622	-0.201

Note: Observations are at the customer level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Cluster-robust standard errors at the market (locality) level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at market level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for consumer-related outcomes family (transaction amount; 0-1 usage of services; 0-1 saved on M-Money) and jointly includes both the pooled and separate treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) in brackets. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates clustered at the locality level.

Firm Revenues (3)

Table C.7: **Effect of information sets on mobile money revenue and business exit**

	Sales (M-Money) per day (GHS)		Business exit indicator	
	(1)	(2)	(3)	(4)
Any treatment	436.6		-0.069	
	(178.4)		(0.058)	
CI: Clustered S.E.	[82.12, 791.1]		[-0.184, 0.046]	
p-value: Wild Bootstrap	0.013		0.286	
p-value: Permutation Test	0.020		0.185	
p-value: R-W MHT Corr (2005)	0.093		0.760	
Lee (2009) Attrition Bounds	<242.9, 486.9>		All obs. selected	
Behaghel et al. (2015) Bounds	<370.1, 467.8>		All obs. selected	
Transparency alone		523.6		-0.069
		(222.0)		(0.058)
CI: Clustered S.E.		[82.44, 964.8]		[-0.184, 0.046]
p-value: Wild Bootstrap		0.027		0.116
p-value: Permutation Test		0.007		0.088
p-value: R-W MHT Corr (2005)		0.715		0.715
Lee (2009) Attrition Bounds		<192.4, 279.7>		All obs. selected
Monitoring alone		418.4		-0.100
		(259.8)		(0.060)
CI: Clustered S.E.		[-96.93, 934.8]		[-0.220, 0.020]
p-value: Wild Bootstrap		0.123		0.151
p-value: Permutation Test		0.000		0.000
p-value: R-W MHT Corr (2005)		0.924		0.354
Lee (2009) Attrition Bounds		<-125.1, 177.8>		All obs. selected
Combined		358.1		-0.017
		(198.1)		(0.076)
CI: Clustered S.E.		[-32.55, 751.8]		[-0.168, 0.132]
p-value: Wild Bootstrap		0.0760		0.819
p-value: Permutation Test		0.000		0.982
p-value: R-W MHT Corr (2005)		0.001		0.001
Lee (2009) Attrition Bounds		<3.990, 148.2>		All obs. selected
Observations	107	107	129	129
Mean of dep var (control)	792.8	792.8	0.218	0.218

Note: Observations are at the select vendor level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Business exits (or deaths) are defined as vendors that were unreachable and/or had inactive registered phone numbers during our endline phone surveys. Heteroskedasticity-robust standard errors are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and heteroskedasticity-robust. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for vendor-related outcomes family (sales revenue; business exit) and jointly includes both the pooled and separate treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) in brackets. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates.

Spillover Effects (4)

Table C.8: **Spillover effects of information sets**

	I. MISCONDUCT FOR UNTREATED BUSINESSES			
	Misconduct indicator		Misconduct amount (GHS)	
	(1)	(2)	(3)	(4)
Any treatment	-0.218		-0.648	
	(0.065)		(0.206)	
CI: Clustered S.E.	[-0.348, -0.088]		[-1.060, -0.235]	
p-value: Wild Bootstrap	0.002		0.002	
p-value: Permutation Test	0.000		0.000	
Lee (2009) Attrition Bounds	<-0.170, -0.155>		<-0.569, -0.479>	
Behaghel et al. (2015) Bounds	NA		NA	
Transparency alone		-0.232		-0.720
		(0.070)		(0.196)
CI: Clustered S.E.		[-0.374 -0.091]		[-1.113, -0.327]
p-value: Wild Bootstrap		0.001		0.000
p-value: Permutation Test		0.000		0.000
Lee (2009) Attrition Bounds		<-0.063, -0.053>		<-0.327, -0.276>
Behaghel et al. (2015) Bounds		NA		NA
Monitoring alone		-0.239		-0.693
		(0.075)		(0.242)
CI: Clustered S.E.		[-0.389, -0.089]		[-1.178, -0.207]
p-value: Wild Bootstrap		0.000		0.001
p-value: Permutation Test		0.000		0.000
Lee (2009) Attrition Bounds		<-0.190, -0.097>		<-0.479, -0.178>
Behaghel et al. (2015) Bounds		NA		NA
Combined		-0.178		-0.524
		(0.070)		(0.224)
CI: Clustered S.E.		[-0.319, -0.037]		[-0.974 -0.075]
p-value: Wild Bootstrap		0.010		0.014
p-value: Permutation Test		0.000		0.000
Lee (2009) Attrition Bounds		<-0.002, 0.091>		<-0.032, 0.345>
Behaghel et al. (2015) Bounds		NA		NA
Observations	411	411	411	411
Mean of dep var in control	0.278	0.278	0.783	0.783

Note: Misconduct variables are audit-based measures. Observations are at non-select vendor x transaction type x transaction date level. Estimations compare non-treated vendors located in treated localities to the pure control localities. Includes randomization strata (district) x transaction type x transaction date dummies. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. 95% confidence intervals are reported in brackets. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates.

Table C.9: **Spillover effects of information sets**

	II. REVENUE FOR NON-MOBILE MONEY			
	Sales (Non M-Money) per day (GHS)		Sales (Total) per day (GHS)	
	(1)	(2)	(3)	(4)
Any treatment	132.7		537.6	
	(58.67)		(195.8)	
CI: Clustered S.E.	[16.19, 249.3.1]		[148.5, 926.7]	
p-value: Wild Bootstrap	0.027		0.008	
p-value: Permutation Test	0.028		0.013	
Lee (2009) Attrition Bounds	<12.61, 103.1>		<172.7, 463.3>	
Behaghel et al. (2015) Bounds	<67.26, 102.8>		<317.0, 440.4>	
Transparency alone		167.1		733.8
		(73.40)		(249.1)
CI: Clustered S.E.		[21.31, 313.0]		[238.7, 1228]
p-value: Wild Bootstrap		0.031		0.008
p-value: Permutation Test		0.004		0.000
Lee (2009) Attrition Bounds		<66.39, 90.84>		<396.7, 523.1>
Behaghel et al. (2015) Bounds		<4.663, 115.7>		<253.2, 611.2>
Monitoring alone		80.51		448.2
		(65.86)		(279.0)
CI: Clustered S.E.		[-50.34, 211.3]		[-106.3, 1002]
p-value: Wild Bootstrap		0.219		0.113
p-value: Permutation Test		0.000		0.000
Lee (2009) Attrition Bounds		<-106.6, -53.65>		<-341.1, 12.15>
Behaghel et al. (2015) Bounds		<-103.4, -54.95>		<-316.2, 2.904>
Combined		141.4		402.5
		(76.43)		(215.7)
CI: Clustered S.E.		[-10.47, 293.2]		[-26.71, 830.7]
p-value: Wild Bootstrap		0.071		0.069
p-value: Permutation Test		0.000		0.000
Lee (2009) Attrition Bounds		<51.71, 101.4>		<-78.58, 96.86>
Behaghel et al. (2015) Bounds		<-24.48, 78.05>		<-174.7, 53.20>
Observations	107	107	107	107
Mean of dep var in control	239.5	239.5	1032	1032

Note: Revenues for Non M-Money are survey-based measures. Observations are at the select vendor level. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Heteroskedasticity-robust standard errors are reported in parenthesis. 95% confidence intervals are reported in brackets. About 75% of vendors bundled M-Money with other business services. Non-mobile money sales code to zero for all outlets that only provide mobile money services. Total sales per day combine mobile money and non-mobile money sales revenues. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates.

Shocks Mitigation (5)

Table C.10: EFFECT OF INFORMATION SETS ON SHOCK MITIGATION

	u-Shocks experience (any)	
	(1)	(2)
Any treatment	-0.068	
	(0.030)	
CI: Clustered S.E.	[-0.128, -0.008]	
p-value: Wild Bootstrap	0.062	
p-value: Permutation Test	0.068	
p-value: R-W MHT Corr ('05)	0.091	
Lee (2009) Attrition Bounds	<-0.078, -0.006>	
Behaghel et al. (2015) Bounds	<-0.073, -0.041>	
Transparency alone		-0.090
		(0.036)
CI: Clustered S.E.		[-0.159, -0.021]
p-value: Wild Bootstrap		0.019
p-value: Permutation Test		0.014
p-value: R-W MHT Corr ('05)		0.033
Lee (2009) Attrition Bounds		<-0.054, -0.007>
Monitoring alone		-0.019
		(0.036)
CI: Clustered S.E.		[-0.088, 0.050]
p-value: Wild Bootstrap		0.707
p-value: Permutation Test		1.00
p-value: R-W MHT Corr ('05)		0.942
Lee (2009) Attrition Bounds		<0.021, 0.043>
Combined		-0.089
		(0.036)
CI: Clustered S.E.		[-0.167, -0.011]
p-value: Wild Bootstrap		0.085
p-value: Permutation Test		0.000
p-value: R-W MHT Corr ('05)		0.091
Lee (2009) Attrition Bounds		<-0.044, -0.018>
Observations	810	810
Mean of dep var (control)	0.895	0.895

Note: u denotes unmitigated and is a 0-1 indicator for whether consumer experienced unexpected shock(s) that s/he could not financially remedy or pay for. Observations are at the customer level. Dependent variable is a survey-based measure. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Cluster-robust standard errors at the market (locality) level are reported in parenthesis. Robustness checks of main results shown for alternative (i) inference and (ii) attrition procedures. Wild bootstrap and permutation test derived from running 1000 replications in each case and clustered at market level. R-W MHT Corr (2005) refers to the multiple hypothesis testing procedure presented in Romano and Wolf (2005) for consumer-related outcomes family (all four 0-1 shock measures; poverty %) and jointly includes both the pooled and separate treatment indicators (i.e., includes both multiple outcomes and multiple tests). 95% confidence intervals (CI) in brackets. Results are similar (i) to exclusion of controls and (ii) to post-double-selection LASSO estimates clustered at the locality level.

C.2 Discussions – Follow-up Surveys with Managers and Vendors in Control Markets

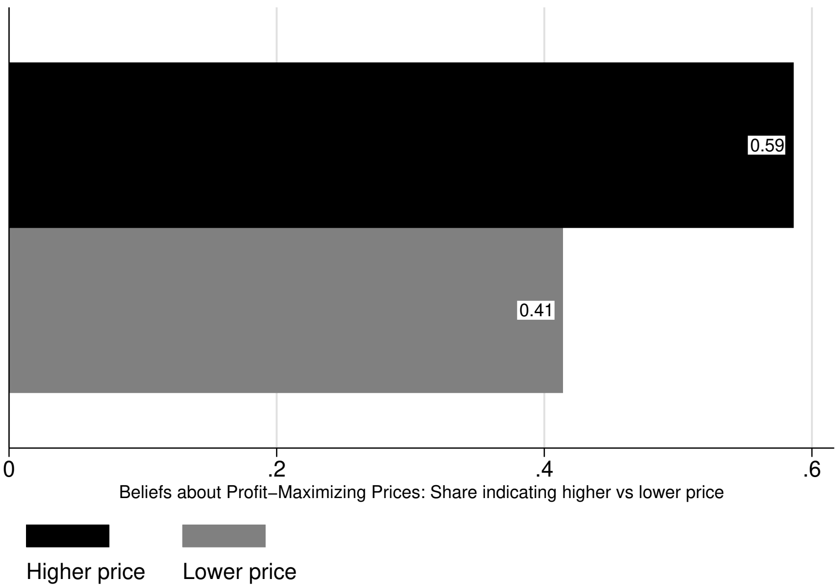
We discuss three implications of our main results.

Implication 1: Why do vendors overcharge – i.e., set higher prices that do not necessarily maximize profits – albeit decrease total welfare? To explore this question, we solicit the beliefs of vendors in control markets about prices. When asked, most of the retailers perceive that higher prices are more profit-maximizing than lower prices (see Figure C.1(a)). We then asked vendors to predict the intervention's likely effect on prices and quantities (treatment effects) [à la DellaVigna and Pope 2018]. Most retailers were incorrect (Figure C.1(b). However, the degree of incorrectness was much larger on quantities than on prices. This makes sense because retailers set prices and the effect on quantities come from consumers response which is possibly related to elasticity. In fact, most retailers predicted effects on prices very well by direction and by trends of the treatments but most predicted the effect on quantities very poorly, particularly by trends of the treatments and effect sizes. This descriptive exercise suggest that vendors commit (unprofitable) misconduct because they perceive that a higher price is better than a lower price. In our context, such perception is reasonable because while vendors can predict very well prices, they cannot predict well the effect on quantities following a price change (i.e., they cannot predict well the price elasticity), leading them to put less weight on quantity effects. To put the predictions into context, vendors predict an overall treatment effect of -36% for prices and +2.7% for quantities, suggesting an elasticity of $\hat{\epsilon} = \frac{2.7\%}{36\%} = 0.075$, which is very small compared to our estimated elasticity of $\epsilon = \frac{45\%}{40\%} = 1.13$. In short, retailers seem to overcharge because of their inability to predict well the price elasticity of demand.

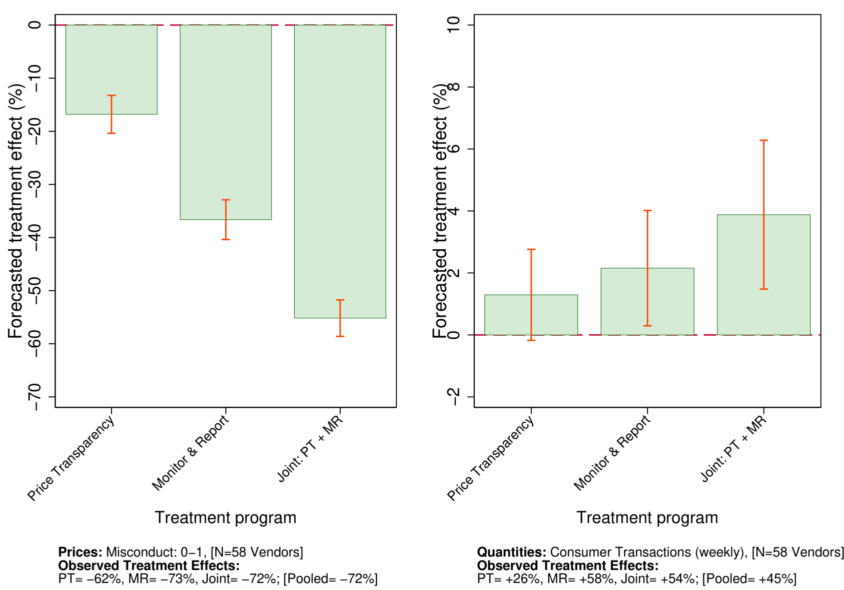
Implication 2: Why did the provider not implement interventions similar to our proposed two-sided information programs, despite the promise of improving provider revenues? To explore this questions, managers – working for the provider – were invited to predict effects of the information interventions on prices and quantities (treatment effects) [à la DellaVigna and Pope 2018]. Most managers were systematically incorrect in their forecasts – incorrectly predicting zero absolute effects for the interventions (Figure C.2(b)). This suggest that the provider did not consider similar interventions because managers (i) were unaware that such interventions and their specific informational contents will work (perhaps, due to the lack of past evidence that these programs work) and (ii) perceive the cost of information campaigns in rural areas to be more expensive (see Figure C.2(a)). We are confident that this will create opportunities for the provider to either take-up or scale-up our two-sided interventions. Previous education to consumers and monitoring efforts by the provider have mainly been in urban areas, tend to focus on large scams, and the few related to retail overcharging are in the form of flyers and text messages, but only few customers can read in rural areas.

Implication 3: Benchmarking the Magnitude of Treatment Effects—The program impacts on quantities are very large. For context, the typical transaction is about GHS100 (based on the audit transactions of GHS50, GHS160 and GHS1100 which were chosen to be typical of the market setting, Table B.9). The regular and official fee will be 1% of this transaction value, which implies a fee value of GHS1.0. The experiment leads the total fee (regular fee + misconduct) to fall from about 1.75% to about 1.10% (Table 2), about a 40% reduction of the transaction fee. The 45% increase in consumer demand (or 52% increase in total vendor sales revenue) in response to a 40% fee reduction is reasonable; it is an elasticity of about 1.13 (or 1.30). In general, consumers are price sensitive to M-Money transactions, as exemplified by recent M-Money taxation and subsidy policy experiments in sub-Saharan Africa (GSMA 2020). In July 2018, the government of Uganda introduced a 1% tax levy on the value of all M-Money transactions (equals 3% charge on transactions) and by August 2018, the overall industry transaction values had decreased 24%; transfer values fell by 50%+; and lower value transactions migrated to cash. Social media tax which was payable via M-Money was imposed around the same time. Similar negative impacts on transactions and business revenues have been observed in Congo (2019 M-Money tax) and Cote d'Ivoire (2018/2019 M-Money tax). When Rwandan providers made transactions "free" in early 2020, M-Money transactions increased to 5 to 6 times their pre-pandemic levels. Overall, in relation to these significant market reactions, which admittedly may conflate a number of issues, our estimated program impacts are very reasonable.

Figure C.1: **WHY MISCONDUCT PREVALENT YET UNPROFITABLE**



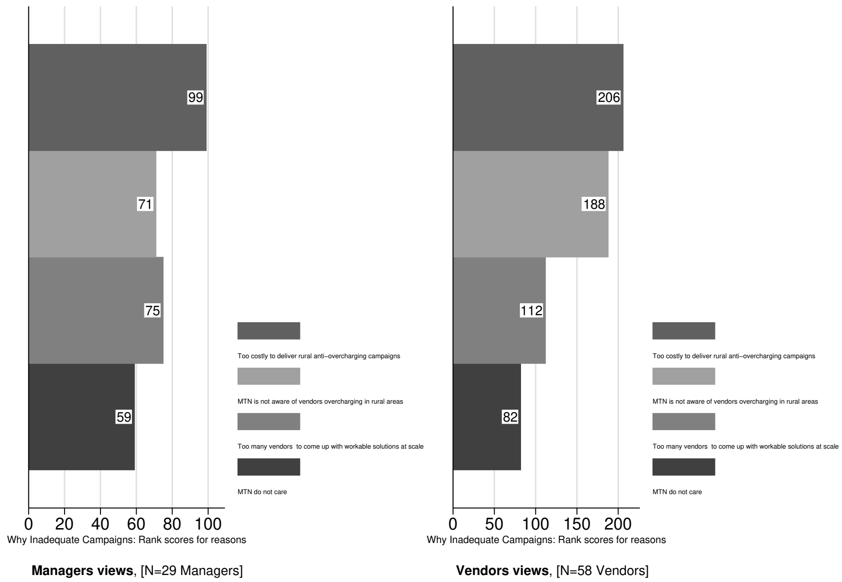
(a) **VENDOR BELIEFS ABOUT PRICES: FOLLOW-UP SURVEYS WITH RETAIL AGENTS**



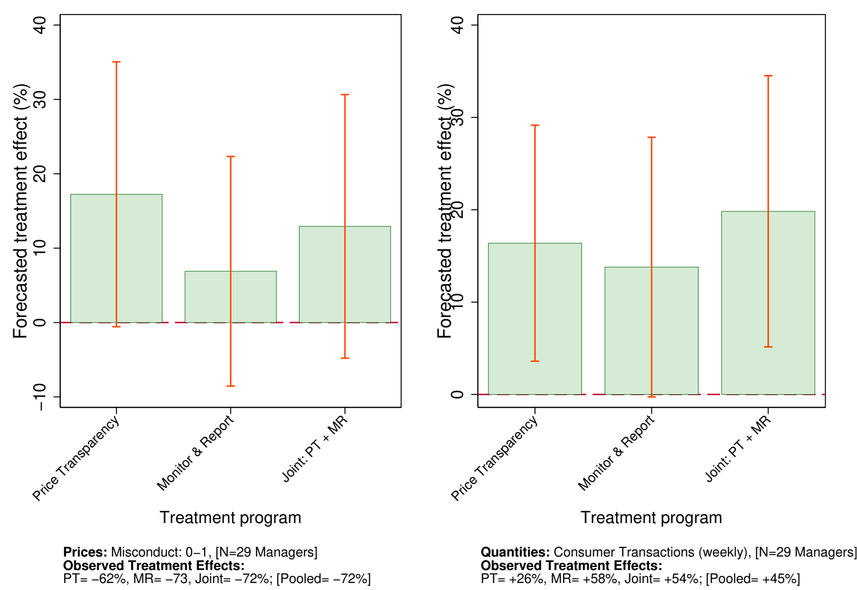
(b) **EVALUATION – VENDOR PREDICTIONS OF TREATMENT EFFECTS**

Note: Figures (top) shows vendors beliefs about profit maximizing prices and (down) shows the predictions of information treatment effects by vendors. (a) Vendors are asked to indicate which price is profit-increasing: 1 (higher=charge above official rates) vs 2 (lower=charge exact official rates). (b) Vendor predictions of treatment effects are shown separately for misconduct/prices (left) and consumers' usage of mobile money/quantities (right). Vendors predict an overall treatment effect of -36% for prices and +2.7% for quantities, suggesting an elasticity of $\hat{\epsilon} = \frac{2.7\%}{36\%} = 0.075$, which is very small compared to our estimated elasticity of $\epsilon = \frac{45\%}{40\%} = 1.13$. The 95% confidence intervals are displayed in the bars of bottom figure. For prices, the predictions are exactly the same for alternative outcomes (misconduct incidence 0/1 vs misconduct amount, GHS). Similarly, for quantities, the predictions are exactly the same for alternative outcomes (consumer transactions per week vs business sales per day).

Figure C.2: **WHY LIMITED PROVIDER INVESTMENTS IN INFORMATION SETS YET PROFITABLE**



(a) **INADEQUATE CAMPAIGNS: FOLLOW-UP SURVEYS WITH MANAGERS (AND RETAIL AGENTS)**



(b) **EVALUATION – MANAGER PREDICTIONS OF TREATMENT EFFECTS**

Note: Figures (top) shows managers and vendors views about why provider's information campaigns have been inadequate and (down) shows the predictions of information treatment effects by managers. Manager predictions of treatment effects are shown separately for misconduct/prices (left) and consumers' usage of mobile money/quantities (right). The 95% confidence intervals are displayed in the bars of bottom figure. For prices, the predictions are exactly the same for alternative outcomes (misconduct incidence 0/1 vs misconduct amount, GHS). Similarly, for quantities, the predictions are exactly the same for alternative outcomes (consumer transactions per week vs business sales per day).

C.3 Heterogeneity

We examine heterogeneity along five dimensions: (i) by vendor competition (market conditions), (ii) seller's gender (market conditions), (iii) by pre-experiment consumer illiteracy (i.e., compliers of the information programs), (iv) by bundled stores, and (v) by beliefs update effect on consumer outcomes (i.e., compliers of the information programs).

First, and motivated by previous theoretical and applied research (Matsa 2011; Annan 2020), we examine heterogeneity by market competition and vendors' gender. Baseline data on vendor sales is used to construct a Herfindahl-Hirschman index, where a lower index reflects higher levels of market competition. The estimates (Tables C.11 and C.12) show that the reduction in misconduct is much larger in localities with more competition,[17] particularly for the combined information program. The effects are similar across gender, which means female vendors might respond more to the information programs because at baseline (pre-treatment), female vendors are significantly more likely to commit misconduct relative to male vendors. This suggests that both underlying market structure and vendors' gender matter for the impact of anti-misconduct information programs. In this case, corrective policies to influence misconduct committed against consumers can include schemes that facilitate information disclosure combined with competition in financial services for the poor, and/or bear on the gender distribution of market vendors.

Second, under much asymmetric information about the true tariffs, consumers might find it difficult to detect, report, and thus reward good vendor behavior, which would be especially true for customers who were vulnerable (i.e., compliers of the information programs: illiterate/ ill-informed, marginalized) at baseline. Tables C.13 and C.14 show consistent evidence. The impact of the intervention on vendor misconduct is larger in markets with high fraction of customers having no formal education at baseline, who also performed poorly in our baseline knowledge tests about the official tariffs.

Third, sellers who operate bundled stores are likely to be more concerned for reputation following our information programs due to relational contracting: vendors can leverage their ongoing customer relationships or goodwill with M-Money transaction services for the other non-M-Money services they provide (Gibbons and Roberts 2012). Thus, we expect the information effects at endline to be larger for vendors who bundle M-Money with other services,

[17] At baseline, vendor misconduct was not significantly different between less and more competitive markets, which can be explained by the existence of much imperfect information. This means some vendors were committing misconduct at baseline, even in markets with more vendor competition, which is consistent with several classic papers discussing the possibility that prices can increase in markets with more firms (see e.g., Satterthwaite 1979; Rosenthal 1980). Consumer search costs can be higher in a larger market with more vendors, which will imply that vendors in larger markets are able to exercise more market power and hence engage in higher misconduct.

relative to market vendors who operate only M-Money services. This is consistent with our earlier evidence indicating large positive spillover impacts of the information program on vendors' non-M-Money sales revenue (Tables 8). Tables C.13 and C.14 show robust evidence that the information effects on misconduct are concentrated on vendors who bundled services.

In addition, to confirm the information effects, we *directly link* the consumer outcomes with measures of belief update (Table C.15). We define update in consumer beliefs about vendor misconduct induced by the information treatment and show strong evidence that the treatment effects on quantities are concentrated in markets where consumers are better calibrated about vendor behavior (i.e., compliers of the information programs).

Table C.11: **HETEROGENEITY BY VENDOR COMPETITION AND BY GENDER**

		MARKET COMPETITION			VENDORS' GENDER	
		Misconduct indicator	Misconduct amount (GHS)		Misconduct indicator	Misconduct amount (GHS)
Any treatment		-0.905	-2.796		-0.254	-0.658
		(0.271)	(1.271)		0.097	(0.295)
		[-1.454, -0.356]	[-5.365, -0.228]		[-0.448, -0.060]	[-1.246, -0.070]
	x Competition	-1.237	-4.303	**x** Female	0.129	0.320
		(0.658)	(2.730)		(0.143)	(0.448)
		[-2.567, 0.093]	[-9.817, 1.211]		[-0.156, 0.415]	[-0.569, 1.211]
	Competition	1.164	3.885	Female	-0.161	-0.396
		(0.655)	(2.817)		(0.131)	(0.434)
		[-0.159, 2.488]	[-0.804, 9.574]		[-0.423, 0.100]	[-1.256, 0.465]
Observations		159	159		335	335
Mean of dep var in control		0.294	0.778		0.294	0.778

Note: Observations are at the select vendor x transaction type x transaction date level. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. 95% confidence intervals are reported in brackets. Market competition index is defined as negative of the Herfindahl-Hirschman (HH) index trimmed to the closed interval $(0, 1)$ to minimize extreme influences, hence the variation in sample sizes.

Table C.12: **HETEROGENEITY BY VENDOR COMPETITION AND BY GENDER**

		MARKET COMPETITION			VENDORS' GENDER	
		Misconduct indicator	Misconduct amount (GHS)		Misconduct indicator	Misconduct amount (GHS)
Transparency alone		-0.652	-2.094		-0.224	-0.549
		(0.321)	(1.565)		(0.109)	(0.326)
		[-1.302, -0.002]	[-5.256, 1.068]		[-0.407, -0.042]	[-1.197, -0.098]
	x Competition	-0.728	-2.802	**x** Female	0.155	0.358
		(0.731)	(3.202)		(0.166)	(0.528)
		[-2.206, 0.749]	[-9.270, 3.665]		[-0.120, 0.432]	[-0.690, 1.408]
Monitoring alone		-0.713	-2.111		-0.237	-0.680
		(0.340)	(1.471)		(0.109)	(0.337)
		[-1.401, -0.025]	[-5.083, 0.860]		[-0.419, -0.054]	[-1.351, -0.010]
	x Competition	-0.742	-2.410	**x** Female	0.086	0.324
		(0.786)	(3.059)		(0.164)	(0.513)
		[-2.329, 0.845]	[-8.589, 3.767]		[-0.186, 0.359]	[-0.695, 1.343]
Combined		-0.965	-2.880		-0.278	-0.673
		(0.291)	(1.333)		(0.104)	(0.317)
		[-1.554, -0.375]	[-5.573, -0.188]		[-0.452, -0.104]	[-1.303, -0.042]
	x Competition	-1.502	-5.028	**x** Female	0.197	0.350
		(0.702)	(2.953)		(0.165)	(0.548)
		[-2.92, -0.084]	[-10.992, 0.936]		[-0.076, 0.472]	[-0.740, 1.440]
	Competition	0.834	2.681	Female	-0.170	-0.407
		(0.704)	(3.068)		(0.134)	(0.440)
		[-0.588, 2.256]	[-3.514, 8.878]		[-0.393, 0.052]	[-1.282, 0.466]
Observations		159	159		335	335
Mean of dep var in control		0.294	0.778		0.294	0.778

Note: Observations are at the select vendor x transaction type x transaction date level. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. 95% confidence intervals are reported in brackets. Market competition index is defined as negative of the Herfindahl-Hirschman (HH) index trimmed to the closed interval $(0, 1)$ to minimize extreme influences, hence the variation in sample sizes.

Table C.13: **HETEROGENEITY BY ILLITERACY AND BY BUNDLING**

		CUSTOMER ILLITERACY [EDUCATION]			BUNDLED STORES	
		Misconduct indicator	Misconduct amount (GHS)		Misconduct indicator	Misconduct amount (GHS)
Any treatment		-0.044	-0.056		-0.138	-0.361
		(0.090)	(0.269)		(0.092)	(0.282)
		[-0.224, 0.135]	[-0.591, 0.477]		[-0.322, 0.045]	[-0.923, 0.199]
	x Illiteracy	-1.139	-3.443	**x** Bundled	-0.350	-0.927
		(0.539)	(1.635)		(0.174)	(0.427)
		[-2.211, -0.066]	[-6.694, -0.193]		[-0.695, -0.004]	[-1.777, -0.078]
	Illiteracy	1.063	3.546	Bundled	0.234	0.654
		(0.524)	(1.730)		(0.162)	(0.378)
		[0.020, 2.106]	[0.108, 6.984]		[-0.089, 0.558]	[-0.096, 1.406]
Observations		332	332		332	332
Mean of dep var in control		0.294	0.778		0.294	0.778

Note: Observations are at the select vendor x transaction type x transaction date level. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. 95% confidence intervals are reported in brackets. Bundled is a 0-1 indicator for whether vendor operates bundled store (M-Money and non-M-Money services). Illiteracy is defined as the market-level fraction of consumers around the vendor that have no formal education at baseline.

Table C.14: **HETEROGENEITY BY ILLITERACY AND BY BUNDLING**

	CUSTOMER ILLITERACY [EDUCATION]			BUNDLED STORES		
		Misconduct indicator	Misconduct amount (GHS)		Misconduct indicator	Misconduct amount (GHS)
Transparency alone		-0.014	-0.084		-0.065	-0.134
		(0.113)	(0.374)		(0.091)	(0.296)
		[-0.239, 0.209]	[-0.829, 0.660]		[-0.247, 0.116]	[-0.724, 0.455]
	x Illiteracy	-1.227	-2.228	**x** Bundled	-0.573	-1.442
		(0.704)	(2.112)		(0.233)	(0.642)
		[-2.626, 0.172]	[-6.426, 1.968]		[-1.037, -0.109]	[-2.719, -0.164]
Monitoring alone		-0.033	-0.016		-0.136	-0.375
		(0.101)	(0.307)		(0.098)	(0.311)
		[-0.234, 0.167]	[-0.627, 0.594]		[-0.331, 0.058]	[-0.993, 0.242]
	x Illiteracy	-1.344	-4.215	**x** Bundled	-0.293	-0.710
		(0.610)	(2.159)		(0.188)	(0.469)
		[-2.557, -0.131]	[-8.505, -0.075]		[-0.668, 0.080]	[-1.643, 0.222]
Combined		-0.076	-0.112		-0.135	-0.338
		(0.101)	(0.324)		(0.098)	(0.309)
		[-0.278, 0.125]	[-0.758, 0.533]		[-0.331, 0.060]	[-0.954, 0.276]
	x Illiteracy	-0.878	-2.875	**x** Bundled	-0.381	-1.078
		(0.635)	(1.625)		(0.178)	(0.454)
		[-2.139, 0.383]	[-6.105, 0.354]		[-0.735, -0.027]	[-1.981, -0.174]
	Illiteracy	1.124	3.720	Bundled	0.250	0.675
		(0.531)	(1.801)		(0.168)	(0.393)
		[0.069, 2.179]	[0.140, 7.299]		[-0.084, 0.585]	[-0.106, 1.456]
Observations		332	332		332	332
Mean of dep var in control		0.294	0.778		0.294	0.778

Note: Observations are at the select vendor x transaction type x transaction date level. Includes (i) randomization strata (district) x transaction type x transaction date dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Cluster-robust standard errors at the vendor level are reported in parenthesis. 95% confidence intervals are reported in brackets. Bundled is a 0-1 indicator for whether vendor operates bundled store (M-Money and non-M-Money services). Illiteracy is defined as the market-level fraction of consumers around the vendor that have no formal education at baseline.

Table C.15: **HETEROGENEITY BY BELIEF UPDATE EFFECT ON QUANTITIES**

	asinh (Total transactions per week)		Used M-Money (last month)	
	(1)	(2)	(3)	(4)
Any treatment x 1(Update)	0.562		0.058	
	(0.194)		(0.026)	
	[0.176, 0.948]		[0.006, 0.110]	
Transparency alone x 1(Update)		0.282		0.014
		(0.309)		(0.041)
		[-0.330, 0.894]		[-0.068, 0.097]
Monitoring alone x 1(Update)		0.732		0.077
		(0.287)		(0.034)
		[0.162, 1.303]		[0.008, 0.145]
Combined x 1(Update)		0.747		0.092
		(0.259)		(0.036)
		[0.233, 1.261]		[0.020, 0.164]
Observations	810	810	810	810
Mean of dep var in control	4.096	4.096	0.734	0.734

Note: Observations are at the customer level. Dependent variables are survey-based measures. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Total transactions per week is the value of M-Money transactions customer conducted in the local market per week at endline. Used M-Money (last month) is a 0-1 indicator for whether the customer used M-Money at endline. "Treatment x **1**(Update)" is the update in customer belief about vendor misconduct induced by the information treatment. The update measure, **1**(Update), is a 0-1 indicator for whether customer's belief agrees with objective audit measure of misconduct for those in the treatment program and equals 0 for those in the control program. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets.

C.4 Alternative Explanations

We discuss potential alternative explanations of the results: (i) Hawthorne effects, (ii) selection effects, (iii) marketing effects, (iv) price and/or (v) bargaining effects.

Hawthorne effects operate in the form of vendors perceiving differential scrutiny by the research teams across treatment arms, and responding to that scrutiny. We mitigate against Hawthorn effects in two main ways. First, our experimental design deploys the research teams to all localities, and so both vendors in the control arm also receive regular visits. Conversely, any additional scrutiny in the treatment arms likely mimics policy-driven scrutiny and thus has high external validity. Second, the very large spillover effects on untreated vendors in treated localities are inconsistent with Hawthorne effects driving the results.

Selection effects operate in the form of exploitative vendors leaving the local market or driving out other vendors *if* business becomes no longer worthwhile through our interventions. The interventions did not lead to significant vendor exits from the local markets (Table 6 and Table B.5), indicating limited selection effects. The programs rather reduced vendor misconduct behavior, which is more consistent with moral hazard and reputation effects (Klein, Lambertz, and Stahl 2016). Thus, when faced with the decision to either exit the market or stay and be honest, vendors choose the latter which is consistent with moral hazard and reputation. Next, marketing/ salience effects operate in the form of our interventions making M-Money services more salient and thus crowding in more customers on the extensive margin, as a result. However, similar to Brown, Hossain, and Morgan (2010), we do not find evidence for an impact on the number of customers (Table C.16), suggesting limited marketing effects.

Price effects can be considered as a by-product of moral hazard and/or reputation effects: vendors take honest actions because of concerns that they might be perceived by consumers as irresponsible, which lead to lower prices and as a result, a price response for consumer demand and other market outcomes. Such price effects are consistent with and re-affirm reputation. We can, however, compare how the treatment effect is driven by (1) the *decrease in prices* against (2) the *decrease in consumers' upwardly-biased beliefs* about misconduct. We separately estimate the treatment effect for consumers that were biased at baseline (reflecting bias + price effects) and for consumers that were unbiased at baseline (reflecting price effect). If we assume that the effects are additive and that the two set of consumers have identical price sensitivity (most of their characteristics are balanced), then the difference between (1) and (2) provides a reduced form estimate of the bias effect. Table C.17 shows the results across different model specifications and suggest that both effects are present but very close in magnitudes (bias effect $\approx$ 55% versus price effect $\approx$ 45% of treatment effect). Next, bargaining effects occur if consumers negotiate with vendors over transaction tariffs. M-Money is not a market where participants negotiate over transactions. By design, the price is ex-ante fixed for a given market transaction and consumers take this as given. Misconduct arises when a vendor decides to overcharge the market transaction. We believe that bargaining is not driving the results.

Table C.16: **MARKETING EFFECTS: EFFECTS ON THE NUMBER OF CUSTOMERS**

	DV: Number of customers (daily)	
	(1)	(2)
Any treatment	-3.422	
	(15.03)	
	[-33.29, 26.44]	
Transparency alone		-4.160
		(15.20)
		[-34.37, 26.05]
Monitoring alone		-7.711
		(16.04)
		[-39.56, 24.17]
Combined		0.972
		(20.17)
		[-39.10, 41.05]
Observations	107	107
Mean of dependent variable (control)	57.44	57.44

Note: Observations are at the select vendor level. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (age, marital status, ethnic group status, employment status, business experience, and bundled store status). Number of customers (daily) is the number of mobile money customers the vendor receives per day at endline. Heteroskedasticity-robust standard errors are reported in parenthesis. 95% confidence intervals are reported in brackets.

Table C.17: **Decomposing price and bias effects on quantities**

	`asinh` (Total transactions per week)				Used M-Money (last month)			
	BIAS=YES		BIAS=NO		BIAS=YES		BIAS=NO	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any treatment	0.901		0.401		0.126		0.055	
	(0.317)		(0.279)		(0.060)		(0.041)	
	[0.263, 1.537]		[-0.155, 0.958]		[0.004, 0.248]		[-0.026, 0.138]	
Transparency alone		0.843		0.190		0.112		0.047
		(0.369)		(0.345)		(0.066)		(0.056)
		[0.102, 1.584]		[-0.498, 0.879]		[-0.201, 0.244]		[-0.065, 0.160]
Monitoring alone		1.135		0.359		0.138		0.030
		(0.438)		(0.339)		(0.070)		(.048)
		[0.255, 2.015]		[-0.316, 1.035]		[-0.002, 0.279]		[-0.065, 0.127]
Combined		0.823		0.660		0.133		0.089
		(0.376)		(0.321)		(0.066)		(0.049)
		[0.068, 1.579]		[0.018, 1.302]		[0.001, 0.265]		[-0.008, 0.187]
Observations	312	312	498	498	301	301	468	468
Mean of dep var in control	4.096	4.096	4.096	4.096	0.734	0.734	0.734	0.734

Note: Observations are at the customer level. Includes (i) randomization strata (district) dummies, (ii) baseline outcomes and (iii) controls (gender, age, marital status, ethnic group status, employment status, education, and income). Total transactions per week is the value of M-Money transactions customer conducted in the local market per week at endline. Used M-Money (last month) is a 0-1 indicator for whether the customer used M-Money at endline. BIAS=YES denotes markets with above median share of biased consumers (biased) at baseline (reflecting bias + price effects). BIAS=NO denotes markets with below median share of biased consumers (unbiased) at baseline (price effects). For each consumer, the bias measure is an indicator for whether customer's subjective belief about misconduct disagrees with objective or administrative audit measure of misconduct. Markets with above median share of biased consumers are classified as biased and those with below median share of biased consumers are classified unbiased. Cluster-robust standard errors at the market (locality) level are reported in parenthesis. 95% confidence intervals are reported in brackets. For `asinh` (Total Transactions per week), bias effect=0.901 (combined effects) - 0.401 (price effect) $\approx$ 0.50 (= 55% of total treatment effect). For Used M-Money, bias effect=0.126 (combined effects) - 0.055 (price effect) $\approx$ 0.07 (= 56% of total treatment effect).

C.5 The Value of Anti-Misconduct Information

The value of information arises from empowering consumers with technologies to enforce market vendors' trustworthy behavior by relying on social sanctions and/or punishment. How cost-effective is our anti-misconduct information intervention? Does this compare well to financial information interventions? When computing cost-effectiveness, we focus on usage of money services-only measure for customers and sales revenue-only measure for vendors. This is a very conservative approach that does not consider the significant treatment effects on savings, risk mitigation outcomes, and other positive externalities, such as increased non-M-Money sales revenue for bundled stores.

We first compute the total cost of interventions to be GHS15,165.[18] With about 730 subjects, we then estimate $\frac{\text{GHS15,165}}{730}$ = GHS20.8 cost per subject, or US$4.0 per person at an exchange rate of US$1=GHS5.12. The opportunity cost of time-use for the subjects is very limited: it takes roughly ten minutes per visit to deliver the information intervention. Compared to Ghana's minimum wage over the period (GHS10.65 per day), the time-use and cost on subjects is very negligible. Thus, the information sets cost approximately US$4.0 per subject. Overall, our cost-effectiveness ratio is 1:5 – a per subject cost of US$4.0 for about +US$19.3 increase in the usage of financial services for customers (Table 5), with sizable implications for consumer welfare (including risk mitigation; see Table 9). This alone suggests a large return of 383%. For vendors, the treatment effect (+GHS437 = +$US85.4; see Table 6) implies a ratio of 1:21 improvement in vendor outcomes.

These rough cost-effectiveness estimates compare favorably with other financial information programs. Frisancho (2018) reports a cost per pupil of US$4.80 and a US$1 increase in financial education program expenses for a 3.3 point improvement in financial literacy. For comparison, we estimate a pooled treatment effect of +27 pp (=+51%) increase in customers' ability to correctly guess seller misconduct behavior. In a recent meta-analysis of financial education interventions, Kaiser et al. (2020) report 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 reduces seller misconduct could be a cost-effective way to improve local markets.

[18] This is based on the number of trained field officers utilized (3 officers), the number of visits to the treated subjects to deliver interventions (3 rounds), transportation costs (GHS385 per officer × 3 officers × 3 rounds = GHS3,465), remuneration and allowance for officers (GHS1,200 per officer × 3 officers × 3 rounds = GHS10,800), and occasional accommodation for officers during field visits (GHS100 per officer × 3 officers × 3 rounds = GHS900). The total cost equals GHS15,165. We reach about 632 panel of treated customers ($\frac{1}{3}\sum_{r=1}^{3}$ number of subjects reachable per round$_r$ = $\frac{629+617+642}{3} = 632$) and about 97 panel of treated vendors ($\frac{98+96+98}{3} = 97$), bringing the total panel number of subjects to 730. Almost all subjects are reached once or twice.

D Information Programs, Data Collection, and Exhibits

D.1 Anti-Misconduct Information Programs

D.1.1 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 understandab
You may call the research team anytime if in any doubt (Phone: XXXXXXXXXX) (omitted to preserve privacy).

Our message is simple. We want to remind you:

- 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: (1) Ask customer to repeat information provided. (2) Ask customer to rate the usefulness of the provided information for their financial decision-making 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.

Our message is simple. We want to remind you:

- If you suspect any discrepancy or glitches in tariffs 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: (1) Ask customer to repeat information provided. (2) Ask customer to rate the usefulness of the provided information for their financial decision-making 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.

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 in tariffs 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: (1) Ask customer to repeat information provided. (2) Ask customer to rate the usefulness of the provided information for their financial decision-making 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.

D.1.2 SECOND: VISIT 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:`

- `They should make sure to ask for official tariff sheets when transacting: e.g., opening new Wallet, OTC, sending.`
- `When opening a new Wallet they should not pay fees - deposit should be credited to their account, they should 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: (1) Ask vendor to rate the usefulness of the provided information for their 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.

- `If they suspect any discrepancy or glitches in tariffs 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: (1) Ask vendor to rate the usefulness of the provided information for their 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.

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, they should 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 in tariffs 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: (1) Ask vendor to rate the usefulness of the provided information for their 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.

D.2 Retail Vendor Points – Photos

Figure D.1: **Vendor banking points**

Note: Providers – MTN Mobile Money, AirtelTigo Money, Voda Cash, GCB Ltd.'s G-Money (new provider)

D.3 Auditors' Training - Measuring Seller Misconduct

INSTRUCTIONS:
VENDOR-BASED APPROVED TRANSACTION TARIFFS

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

KEY: TRANSACTIONAL CODES
OVER-THE-COUNTER, OTC

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

TOKEN

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

FALSIFY [INSTANT VERIFIABILITY PROVIDED BY PROVIDER]

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

ACCOUNT OPENING

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

TRANSACTION APPROACH

**DURING VISIT (Very simple language, no deviations allowed): `Good morning/afternoon/evening.`

`I want to make a M-Money transaction [USE CODES: T1...T12].`

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

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

ADDITIONAL NOTES

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

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

Q0	Q1a	Q1b	Q1c	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
No.	VISIT DATE			Locality code?	"Rep" Vendor code?	TRANSACTION TYPE? USE CODES: T1...T12	Transaction OVERCHARGED? 1=YES; 2=NO=>Q7	How much DIFFERENCE? GHS	Transaction successful? 1=YES; 2=NO; 3=NO CASH	Appx wait time transaction took? MINS	Related to Vendor just visited? 1=YES; 2=NO => Q11	How are you related to Vendor? 1=RELATIVE; 2=FRIEND; 3=OTHER	Vendor's Gender? 1=MALE 2=FEMALE	Vendor involved in non-Mobile Money businesses? 1=YES 2=NO	Tariffs posted? 1=YES; 2=NO
	MM	DD	TIME												
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
.															
.															
.															
.															
.															
.															

. . . . . .

D.4 Definition of Relevant Select Variables – Questions

Consumer outcomes:

1. Uptake of transactional services:
   (a) `Consider the last month - What is the typical value of Mobile Money transactions (cash-in and cash-out) you conducted in this locality per week? (NOTE 1: Please only include transactional estimates in seven (7) days. NOTE 2: Ask the customer to refer to his or her records/diaries for past days in case forgotten)` **`GHS/week...`**
   (b) 0-1 Indicator for whether consumer used M-Money (last month): `If` **`1(a) > 0GHS`**
2. Savings likelihood:
   (a) `Consider the last month - From a scale of 1 (low) to 5 (high), how likely are you to save generally on M-Money now?` **`1=Very low, 2=Low, 3=Medium, 4=High, 5=Very high`**
   (b) 0-1 Indicator for whether consumer saved on M-Money (last month): `If uptake if` **`2(a) > 2 (median)`**
3. Unexpected shocks mitigation: `Have you experienced any of the following common shocks within the past three (3) months where you or your household did not have enough cash or M-Money resources on hand to cover costs?` **`0=No, 1=Yes`**
   (a) `Death of a close person, relative, or friend? (death)`
   (b) `Unexpected loss of revenue or wages, e.g., via unemployment, bad business? (revenue)`
   (c) `Unexpected illness, accident, or health condition? (health)`
   (d) `Any general floods or droughts? (weather)`
   (e) `Unexpected (high/low) food prices? (prices)`
   (f) `Other unexpected shocks (i.e., weather, input prices, diseases) that affect your farm production or house expenses? (house expenses)`
   (g) Shocks experience = 1 `if any of` **`3(a) - 3 (f) = Yes`**
4. Subjective beliefs (perception about seller misconduct) - Highly correlated responses:

(a) Non-incentivized beliefs statement: `Consider the past four (4) months (interventions in force) - In my [research enumerator's] view, M-Money vendors generally overcharge customers' transactions at vendor points (seller misconduct).` **`1=Agree, 2=Disagree`**

(b) Incentivized beliefs: `What's your [customer's] estimate of the % of others (all vendors and customers in this locality) that will Agree with statement 4(a)?` **`%...`**

- NOTE: Customers are jointly asked to guess the percentage of others (all vendors and customers in their locality) who would `Agree` with statement 4(a) (beliefs about others' beliefs). To incentivize their reports, among all respondents in a locality, the respondent with the closest guess (to the locality-level estimate) immediately receives 10GHS after all respondents have answered, either in cash to their M-Money accounts or in-kind through a phone calling-credit. All respondents are informed of this payoff before answering.

(c) `Consider the past four (4) months (interventions in force) - Any experiences of overcharged M-Money fees at Mobile Money centers?` **`0=No, 1=Yes`**

Business outcomes:

1. Sales revenue (Mobile Money): `Consider the last month - What was the total sales the Mobile Money business made daily? (NOTE 1: think about all cash-in and cash-out transaction volume records. NOTE 2: Ask the vendor to refer to his or her records/diaries for past days in case forgotten)` **`GHS/day...`**

2. Sales revenue (non-Mobile Money): `What was the total sales the non-M-Money business made daily considering the last month (NOTE 2: Ask the vendor to refer to his or her records/diaries for past days in case forgotten)?` **`GHS/day...`**

3. Total sales revenue = `1+2,` **`GHS/day...`**

Control set:

1. Bundling (bundled stores): `Currently do you [vendor] offer other services at your business center, other than M-Money? Example - sell provisions, airtime, phones, accessories, appliances, etc.` **`0=No, 1=Yes`** `(Alternative measure: see Q12 in Table D.1)`

2. Tariff posting: `Consider the last thirty (30) days or last month: How often do you [vendor] post your tariff sheets at your banking point in a typical week?` **`1=Never (less than 1 time in 7 days), 2=Sometimes (1-2 times in 7 days), 3=Often (3-4 times in 7 days), 4=Very often (5-7 times in 7 days)`** `(Alternative measure: see Q13 in Table D.1)`
3. Age: `What is your [vendor/customer] age?` **`Years`**
4. Married: `Are you [vendor/customer] married?` **`0=No, 1=Yes`**
5. Akan: `What is ethnicity do you [vendor/customer] identify with?` **`1=Akan, 2=Ewe, 3=Ga-Dangme, 4=Others`**
6. Self-employed: `Are you [vendor/customer] self-employed?` **`0=No, 1=Yes`**
7. Business experience: `How long have you been in the Mobile Money service business?` **`Years`**

Poverty Scorecard (Schreiner 2005):

1. `How many members does the household have?` **`Use codes: 0=Eight+ 4=Seven 9=Six 13=Five 14=Four 21=Three 24=Two 29=One`**
2. `Are all household members ages 5 to 17 currently in school?` **`0=No 2=Yes 3=No one ages 5 to 17`**
3. `Can the male head/spouse read a phrase/sentence in English?` **`0=No 2=No male head/spouse 5=Yes`**
4. `What is the main construction material used for the outer wall?` **`0=Mud bricks/earth, wood, bamboo, metal sheet/slate/asbestos, palm leaves/thatch (grass/raffia), or other 5=Cement/concrete blocks, landcrete, stone, or burnt bricks`**
5. `What type of toilet facility is usually used by the household?` **`0=No toilet facility (bush, beach), or other 4=Pit latrine, bucket/pan 4=Public toilet (e.g., WC, KVIP, pit pan) 6=KVIP or WC`**
6. `What is the main fuel used by the household for cooking?` **`0=None, no cooking 6=Wood, crop residue, sawdust, animal waste, or other 13=Charcoal, or kerosene 22=Gas or electricity`**

7. Does any household member own a working box iron or electric iron? **0=No 4=Yes**

8. Does any household member own a working television, video player, VCD/DVD/MP3/MP4 player/iPod, or satellite dish? **0=No 2=Only television 3=Video player, VCD/DVD/MP3/MP4 player/iPod, or satellite dish (regardless of T.V.)**

9. How many working mobile phones are owned by members of the household? **0=None 4=One 8=Two 10=Three+**

10. Does any household member own a working bicycle, motor cycle, or car? **0=No 3=Only bicycle 8=Motor cycle or car (regardless of bicycle)**

D.5 Follow-up Surveys – Design and Questions

Managers:

- Respondents: Provider (MTN M-Money+): Compliance department Managers + Staff, Managers in HQ, Managers in Eastern Regional Office, Managers in 9 District Offices in the study area [West Akim, East Akim, Suhum Kraboa, Nsawam Adoagyir, New Juaben, Akwapim North, Yilo Krobo, Lower Manya, Asuogyaman]
- Welcome and thank-you for answering this brief research survey about the market for M-Money.

  - q1. Are you aware of vendors' overcharging MOMO transactions? **1=yes/ 2=no**
  - q2. If yes, how? **1=experienced it/ 2=handled consumer reports/ 3=heard of it/ 4=other, specify, and indicate response**
  - q3. Which of the ff. would you say is the most important issue to you and your department (as a measure of your department's performance)? **1=fraud issues/ 2=technical issues/ 3=marketing issues/ 4= other, specify, and indicate response**

- **EXERCISE 1: LEXICOGRAPHIC RANKINGS**
- QL1. Please rank the following 9-hypotheses in order of the most plausible (1 being most plausible) reason for "why vendor misconduct or overcharging on MTN MOMO is prevalent in low-income areas?" — based on your view

  - Limited investment in campaigns and customer service in "rural areas" by the Provider -> Two ways (either): Inadequate-investment: MTNs education tends to focus on large scams and the little on "overcharging" is mostly in the form of flyers/texts but few customers do read or can even read such vs Under-investment: all redress channels are centralized in capital cities, and so not locally or low-income area friendly

  - Poorly informed consumers about prices (and about how to report or easiness to report overcharging if any) - This also separately exacerbates the effects from other reasons
  - Vendor beliefs about "profit-maximizing" charges or prices are possibly misguided
  - Low vendor commissions (or business income) so, overcharging is a short-run incentive to rip off consumers
  - Lack of competition and /or limited substitution -> Two ways (either): (i) Few vendors to switch to vs (ii) Few Banks available to switch to in some markets
  - The [perceived] cost of overcharging is low to the vendors, especially in environments where misconduct is hard to detect or hard to report, etc.
  - Weak agency relationship between vendors and Provider: Vendors don't care if their actions harm Provider because they think the Provider doesn't care much about them.
  - Limited consumer search behavior: consumers hardly switch to use different vendors other than the nearby vendor they are used to, which undermine any possible competition effects. aka. "Ghanaians don't like change"
  - Heterogeneity: some agents are made better off and others worse off, such that there is no average/overall impact

- QL2. Please rank the following 5-hypotheses in order of the most plausible (1 being most plausible) reason for "why MTN has not fully solved the problem of their vendors overcharging in low-income areas (yet if the vendors overcharge transactions it can harm MTN's total revenues or total transactions)" — based on your view

  - Too costly to deliver anti-overcharging campaigns and services, particularly in very low-income areas (this may explain why currently all MTN offices and redress channels are in national or district capitals)
  - MTN is not aware of vendors overcharging in rural areas
  - Scale effects: Too many vendors (because vendor entry into the market is easy) such that the Provider, who has over 90%+ market share, now cannot even handle them. I.e., MTN finds it difficult to come up with workable solutions at that scale
  - MTN don't care. I.e., MOMO market is almost 90% dominated by MTN, so it has monopoly power and perhaps gives it a lax attitude to deal with the issue
  - Other(s) major reason, please specify, and indicate rank

- **EXERCISE 2: PREDICTING TREATMENT EFFECTS**

- QT1. Treatment scenario #1: We consider a simple information intervention across low-income communities that **"first provides consumers with education about official MOMO transaction charges, and second then alerting the nearby vendors that customers have been given such training or market information".**

- QT2 . Treatment scenario #2: We consider a simple information intervention across low-income communities that **"first provides consumers with education about MTN's toll-free number to report any suspected overcharges, and second then alerting the nearby vendors that customers have been given such training or market information".**

- QT3. Treatment scenario #3: `We consider a simple information intervention across low-income communities that` **`"first provides consumers with education (i) about official MOMO transaction charges and (ii) about MTN's toll-free number to report any suspected overcharges, and second then alerting the nearby vendors that customers have been given such training or market information".`**

  - For each of QT1 (PT-alone), QT2 ((MR-alone), QT3 (PT+MR): What will be your estimate of how this information program will affect:

    - Prices: i.e., The overall probability of nearby vendors' overcharging (yes/no) will change by? `1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ...`
    - Prices (value, GHS): i.e., The overall overcharged amount or value (GHS) by nearby vendors' will change by? `1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ...`
    - Quantities (demand): i.e., Consumers' usage of M-Money services (weekly) will change by? `1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ..`
    - Quantities (sales): i.e., Daily transaction amount at the nearby vendor points will change by? `1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ...`

Vendors:

- Respondents: `Vendors in control markets`
- Welcome and thank-you for answering this brief research survey about the market for M-Money.

  - `q1. Are you aware of vendors' overcharging MOMO transactions? 1=yes/ 2=no`
  - `q2. If yes, how? 1=experienced it/ 2=handled consumer reports/ 3=heard of it/ 4=other, specify, and indicate response`
  - `q3. Which of the ff. would you say is the most important issue to you and your business (as a measure of your business's performance)? 1=fraud issues/ 2=technical issues/ 3=marketing issues/ 4= other, specify, and indicate response`

- **EXERCISE 1: LEXICOGRAPHIC RANKINGS**
- QL1. Please rank the following 9-hypotheses in order of the most plausible (1 being most plausible) reason for "why vendor misconduct or overcharging on MTN MOMO is prevalent in low-income areas?" — based on your view

  - Limited investment in campaigns and customer service in "rural areas" by the Provider -> Two ways (either): Inadequate-investment: MTNs education tends to focus on large scams and the little on "overcharging" is mostly in the form of flyers/texts but few customers do read or can even read such vs Under-investment: all redress channels are centralized in capital cities, and so not locally or low-income area friendly
  - Poorly informed consumers about prices (and about how to report or easiness to report overcharging if any) - This also separately exacerbates the effects from other reasons
  - Vendor beliefs about "profit-maximizing" charges or prices are possibly misguided
  - Low vendor commissions (or business income) so, overcharging is a short-run incentive to rip off consumers
  - Lack of competition and /or limited substitution -> Two ways (either): (i) Few vendors to switch to vs (ii) Few Banks available to switch to in some markets
  - The [perceived] cost of overcharging is low to the vendors, especially in environments where misconduct is hard to detect or hard to report, etc.
  - Weak agency relationship between vendors and Provider: Vendors don't care if their actions harm Provider because they think the Provider doesn't care much about them.
  - Limited consumer search behavior: consumers hardly switch to use different vendors other than the nearby vendor they are used to, which undermine any possible competition effects. aka. "Ghanaians don't like change"
  - Heterogeneity: some agents are made better off and others worse off, such that there is no average/overall impact

- QL2. Please rank the following 5-hypotheses in order of the most plausible (1 being most plausible) reason for "why MTN has not fully solved the problem of their vendors overcharging in low-income areas (yet if the vendors overcharge transactions it can harm MTN's total revenues or total transactions)" — based on your view

  - Too costly to deliver anti-overcharging campaigns and services, particularly in very low-income areas (this may explain why currently all MTN offices and redress channels are in national or district capitals)
  - MTN is not aware of vendors overcharging in rural areas
  - Scale effects: Too many vendors (because vendor entry into the market is easy) such that the Provider, who has over 90%+ market share, now cannot even handle them. I.e., MTN finds it difficult to come up with workable solutions at that scale
  - MTN don't care. I.e., MOMO market is almost 90% dominated by MTN, so it has monopoly power and perhaps gives it a lax attitude to deal with the issue
  - Other(s) major reason, please specify, and indicate rank

- **EXERCISE 2: PRICING BELIEFS AND PREDICTING TREATMENT EFFECTS**
- **PRICING BELIEFS:**

- QP1. In your view, which of the following price actions do you think is more profit-maximizing 1=charging above the official price (I.e., high charges paid but faced with fewer customers in return) vs 2=charging the exact official price (I.e., low charges paid but faced with more customers in return)
- QP2. In your view, if you overcharge, it affects 1=only your transactions and profits vs 2=only your rivals transactions and profits vs 3=both you and your rivals' transactions and profits
- QP3. In your view: if your rival overcharges, it affects 1=only your rivals transactions and profits vs 2=only your transactions and profits vs 3=both your rival and your transactions and profits

- **PREDICTING TREATMENT EFFECTS:**
- QT1. Treatment scenario #1: We consider a simple information intervention across low-income communities that **"first provides consumers with education about official MOMO transaction charges, and second then alerting the nearby vendors that customers have been given such training or market information".**
- QT2. Treatment scenario #2: We consider a simple information intervention across low-income communities that **"first provides consumers with education about MTN's toll-free number to report any suspected overcharges, and second then alerting the nearby vendors that customers have been given such training or market information".**
- QT3. Treatment scenario #3: We consider a simple information intervention across low-income communities that **"first provides consumers with education (i) about official MOMO transaction charges and (ii) about MTN's toll-free number to report any suspected overcharges, and second then alerting the nearby vendors that customers have been given such training or market information".**

  - For each of QT1 (PT-alone), QT2 ((MR-alone), QT3 (PT+MR): What will be your estimate of how this information program will affect the ff:
    - Prices: i.e., The overall probability of nearby vendors' overcharging (yes/no) will change by? 1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ...
    - Prices (value, GHS): i.e., The overall overcharged amount or value (GHS) by nearby vendors' will change by? 1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ...
    - Quantities (demand): i.e., Consumers' usage of M-Money services (weekly) will change by? 1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ..
    - Quantities (sales): i.e., Daily transaction amount at the nearby vendor points will change by? 1= -100% 2= -75% 3= -50% 4= -25% 5= -/+0 (unaffected) 6= +25% 7= +50% 8= +75% 9= +100% ANSWER: ...