

The Good Wife? Reputation Dynamics and Financial Decision-Making Inside the Household*

Nina Buchmann
Stanford University

Pascaline Dupas
Stanford University and NBER

Roberta Ziparo
Aix-Marseille School of Economics

August 11, 2021

Abstract

We study the dynamic relationship between intra-household specialization, resource allocation and women's investment decisions. We consider public good investments delegated to the wife in settings where wives perceived to be savvy investors by their husbands are entrusted with a larger share of the budget. We show, first theoretically, then empirically through a series of experiments with couples in Malawi, that a signaling game can result, in which wives, in order to maintain control over a larger share of the budget, (a) under-invest in novel goods with unknown returns; and (b) knowingly over-use low-return goods in order to hide bad purchase decisions—we call this the *intra-household sunk cost fallacy*. These dynamics matter for women's well-being as well as for the design of poverty alleviation programs.

Keywords: Intra-household model, Signaling, dynamic games, experiment, technology adoption. JEL code: D13, J16, O12.

*We are immensely grateful to Aaron Wolf for outstanding field research assistance, and to all our study participants for their time and attention. The data collection was done with the assistance of IKI Malawi. The research protocols were approved by the Stanford IRB and the Malawi National Committee on Research in the Social Sciences and Humanities. The experimental designs were pre-registered in the AEA RCT registry (ID 4069). We gratefully acknowledge funding from Stanford University. We thank Renadu Bourlès, Lukas Bolte, Marcel Fafchamps, Yucheng Liang, Madeline McKelway, Muriel Niederle, Agathe Pernoud, Collin Raymond, Alessandra Voena, and seminar audiences at Stanford's Eddie's Lunch, Stanford's Gender Coffee, THreD, Penn State, UCL, Bristol and the NYU BREAD 2021 conference for very helpful comments. All errors are our own.

1 Introduction

Many women worldwide rely on transfers from their spouse for general household expenditure. This is especially true for women in poor countries, where women’s earnings are limited by a number of factors, including labor-intensive home production (Jayachandran, 2015). What drives the size of such transfers and the discretion women have over how to spend them? Existing household models and their empirical tests have documented the role of outside options. What if a woman’s access to household resources also depends on her spouse’s beliefs about her ability to manage such resources?

This paper brings to light reputation dynamics within the household. The key intuition is the following. If the husband’s transfer to the wife is a function of his belief about her savviness at using household resources, the wife has an incentive to manipulate information about her type, and such strategic behavior may create inefficiencies. In particular, the wife will want to: i) under-invest in risky but potentially high-return goods, to avoid non-savvy purchases of unproductive goods (“lemons”); or ii) exert costly effort to hide that she has purchased a lemon.

We develop a signaling model to formalize this intuition and generate a series of testable predictions. We then present results from a suite of experiments involving over 2,000 married individuals in rural Malawi to test the model. The experimental results are consistent with our theoretical predictions, and suggest that reputation concerns substantially affect women’s decision-making. In particular, they limit women’s willingness to try out new technologies with unknown but potentially high returns for the household. These findings have important implications for understanding women’s agency inside the household and their well-being, as well as for the design of effective anti-poverty programs.

We propose a repeated signaling game in which the husband decides each period whether to delegate a share of the discretionary household budget to the wife, and the wife decides what public goods to spend it on: a safe good with known returns, or a risky good that can be either high-return or zero-return (e.g., a new cookstove advertised as more efficient or a new grain supposed to be drought resistant). Wives vary in their “expertise” at assessing the productivity of risky goods (think savvy vs. gullible): expert wives are better at assessing the productivity of risky goods and less likely to invest in risky goods that are zero-return. The husband is better off delegating to expert wives only. However, he does not directly observe the productivity of the investments, so he cannot directly detect errors. Instead, he learns about his wife’s type by observing the rate at which she invests in risky goods, as well as whether she uses the goods she invests in.

We show that two types of inefficiencies may arise: under-investment in potentially high-return goods and over-use of unproductive goods (lemons). Their relative importance depends on the cost of hiding the purchase of a lemon. If the cost is low (e.g., it is possible to pretend a new stove is

efficient by hiding how much firewood is collected and used), a pooling equilibrium will emerge: non-expert wives' investment probability will converge towards the investment rate of expert wives. Given their lack of expertise they commit both type II errors (missing out on high-return goods) and type I errors (purchase of lemons), both of which are unobserved by the husband. Thus, the first inefficiency is that non-expert wives do not experiment with new goods as much as they should, because experimenting too much would reveal their type. The second inefficiency is that non-expert wives have to overuse lemons to “save face”: they act sub-optimally after a type I error (the purchase of a lemon) to hide their mistake and protect their reputation. This generates (and rationalizes) a behavior empirically equivalent to the sunk cost fallacy. In contrast, if the cost of hiding the purchase of a lemon is high (e.g., it is quasi-impossible to hide that a crop did not withstand drought), a separating equilibrium will emerge, in which non-expert wives shy away from investing in goods with uncertain returns. In this case, there is no overuse of lemons, but the under-investment caused by reputation concerns becomes very severe: in order to commit no type I error, non-experts forgo investment almost completely.

We run a series of experiments in Southern Malawi to test the model.¹ We start by running two parallel lab-in-the-field experiments, one with husbands and one with wives, with about 1,100 married couples randomly sampled from 36 villages. The experiment with husbands (the “transfer experiment”) is designed to test whether reputation matters for the share of the discretionary household budget that husbands delegate to their wives. We play a dictator game with multiplier between the husband and his wife. We randomly assign husbands to a “salience” treatment: right before the transfer decision, husbands in the salience treatment are asked to recall examples of their wife’s (potential lack of) market expertise. Husbands in the control group are asked the same questions only *after* the transfer decision has been made. We find that the salience treatment substantially decreases transfers to women perceived as less savvy. This is consistent with the model’s premise that husbands’ transfers to their wives depend on their beliefs about their wife’s expertise.

The experiment with wives (the “signaling experiment”) is designed to test whether women’s decisions are influenced by reputational incentives. For this, we offer wives the opportunity to participate, with financial compensation, in a quiz in which they assess the quality of a series of market goods (e.g., natural sponge vs. plastic sponge), and their performance on the quiz is shared with their husbands. We inform wives that if they choose to participate, they will be allowed to use part or all of their survey compensation to correct mistakes, i.e., to improve their score before the signal is sent to the husband. In the terms of the model, this is akin to hiding the purchase of a lemon. We randomize whether the hiding cost (the cost per corrected answer) is low or high. Consistent with the model, we find that wives facing a greater risk of error (“non-experts”) participate as

¹All experiments were pre-registered in the AEA RCT Registry (#4069).

much as expert wives when the cost of hiding is low (pooling), but significantly less when the cost of hiding is high (separation). In addition, conditional on participating, non-expert wives hide significantly more than expert wives. As a result, the signals sent to the husbands of women who choose to participate are identical for expert and non-expert wives, even though initial scores are lower among non-expert wives. On average, wives forgo MWK 103 in experimental earnings in order to avoid a potentially bad signal—about 30% of average daily income. 28% do not participate at all (they forgo the participation fee so no signal is sent, i.e., their husbands never find out about the game) and 20% participate but use the participation fee to hide mistakes.

The third experiment (the “market experiment”) is designed to test whether real-life investment decisions by women are influenced by reputational concerns and whether the difference in investment rates between experts and non-expert wives decreases as uncertainty about the quality of the risky good decreases. For this, we run a field experiment with 675 married women sampled while shopping (alone) at one of six large markets. We administer a survey, pay a compensation fee, then elicit women’s willingness to exchange some of their compensation fee for an unfamiliar good with high returns, which we explain to the women using scientific evidence. We experimentally vary whether the husband will know that the good is high-return, and whether the husband will know that the cost of the good is lower than its payoff. Consistent with the model, low-expertise women have a 22% lower willingness to pay to purchase the good than women with high expertise. This gap disappears when the husband receives information that removes the reputation risk from the wife’s investment decision (pooling).

Together, our experimental results are consistent with the intuition that women’s reputation within the household matters. Transfers from husbands are influenced by it, and women forgo substantial payments as well as valuable investments when doing so helps maintain their reputation inside the household. Importantly, in all three experiments, there is heterogeneity based on how much wives received in real-live transfers from their husbands in the preceding two months. The patterns predicted by the model are observed only in couples in which monthly transfers from the husband to the wife are high (controlling for household income). This is consistent with the model: couples with high transfers are those in which the wife’s reputation is still above the cutoff reputation above which discretionary transfers occur, i.e., couples where reputation dynamics are still at play. This suggests that the behavior of both spouses in the experiments is most likely driven by real-life concerns about resource allocation—not experimenter demand effects.

Our empirical evidence comes from a specific setting, but the mechanism in the model could be at play in rich and poor households alike. For instance, a bay area husband might continue using the yogurt maker that produces tasteless yogurt in order to maintain his reputation as an investor—and be able to buy a Sourdough bread maker in the future.² The model could also apply to parents

²No identification with actual persons is intended or should be inferred.

and teenage children, or international migrants and those they send remittances to. We see the dynamics we model as particularly consequential for poor households in low-income countries, however, where the subordinate position of one spouse with respect to the other is still common and exacerbates the issue.

The intra-household dynamics we bring to light theoretically and empirically may be one of the factors behind the low take-up of new technologies targeted towards women, such as preventative health products (Cohen and Dupas, 2010; Meredith et al., 2013), improved cookstoves (Mobarak et al., 2012), etc. Understanding the extent to which intra-household dynamics influence the ability of women to experiment with new technologies is an important step towards understanding the types of policies and programs that can influence adoption. In particular, our findings suggest that marketing campaigns promoting new technologies specifically to women, as many non-governmental organizations do, may generate negative consequences for women who face reputational risks when asked to make investment decisions on behalf of their household.

Our study helps to uncover some implications of reputation concerns in the management of the resources of the household. As such, it connects the literature on bargaining within the household (Manser and Brown, 1980; Chiappori, 1988, 1992) with the literature on dynamic signaling (Noldeke and van Damme, 1990; Swinkels, 1999; Kremer and Skrzypacz, 2007). So far, most economic models of the household have focused on identifying consumption shares: how much is spent on public vs. private consumption, and the relative consumption shares of the husband and the wife. There is no asymmetry of information, and shares are determined by the outside options of the two parties (Manser and Brown, 1980; Chiappori, 1988, 1992). Our contribution is to study the impact of information asymmetries about the wife's expertise.

The paper also contributes to the intra-household economic literature showing the existence of strategic behavior inside the household. While inefficient management of financial resources (Anderson and Baland, 2002; Ashraf, 2009; Schaner, 2015), hiding (Ashraf et al., 2014; Boltz et al., 2016), lying (Ashraf et al., 2020; Apedo-Amah et al., 2020) and strategic use of domestic violence (Tauchen et al., 1991) have been documented, less is known about reputation dynamics inside the household. Reputation concerns within the family have previously been proposed in the interaction between parents and children: parents can have a strategic incentive to act "tough" with older children in order to build a reputation as non-lenient, and thereby dissuade later-born children from misbehaving, e.g., studying too little (Hao et al., 2008; Fu and Pantano, 2015; Hotz and Pantano, 2015). We bring to the fore the fact that reputation mechanisms between spouses can affect decision-making and behavior, and may matter for the design of social policies.

In addition, the model rationalizes, within the realm of neoclassical economic mechanisms, i.e., without the need for any psychological factor or behavioral bias, a behavior that looks empirically akin to the sunk cost fallacy—the greater tendency to continue an endeavor once an investment in

money, effort, or time has been made (Arkes and Blumer, 1985). Under standard economic models this behavior is irrational because, once the expense has been incurred, it should be irrelevant to the decision to go on or not. The idea that such behavior may be driven by the rational need to save face when future payouts are at stake has been previously modeled in the context of firm managers by Kanodia et al. (1989) and Prendergast and Stole (1996). To our knowledge, we are the first paper to derive the existence of such a phenomenon in a setting where the cost of hiding private information is borne by the agent (the wife)—while in the firm setting the escalation costs are borne by the principal (the firm).

The remainder of this paper is structured as follows. Section 2 presents the model and derives a set of testable predictions. Section 3 presents the empirical setting. Sections 4, 5, and 6 describe the transfer experiment, the signaling experiment and the market experiment, respectively, each testing different theoretical premises or predictions. Section 7 tests for heterogeneity by the size of the discretionary transfers from the husband to the wife, and section 8 concludes.

2 A Signaling Model with Endogenous Budget Allocations

We model the effect of reputation concerns on women’s investment decisions and hiding of investment mistakes. We consider a set-up with two spouses (we call them “husband” and “wife”), who play an infinite horizon game.³ To study the role of reputation concerns in the intra-household sharing of resources, we assume that the outside option of both spouses is constant.⁴

Game Structure The two spouses play a dynamic game. The structure of the game is common knowledge to both spouses. In each period, the structure is as follows:

1. Husbands decide whether to make financial transfers to their wives.
2. Wives – that can be of two types: expert and non-expert – decide whether to invest in a risky good with unknown but potentially high return or a safe good with certain but low return. Expert wives learn the productivity of the risky good before making their investment choice while non-expert wives do not.
3. After buying the good, all wives observe the productivity of the good they purchased and decide whether to use the good.

³This is a common assumption in games that are played so frequently that the horizon approaches only very slowly and is thus ignored, i.e., it does not enter people’s strategic calculations (Mailath and Samuelson (2007), section 4.1).

⁴This assumption implies that the bargaining power of the two spouses is constant apart from the reputation dynamics analyzed in the paper. Furthermore, we set up the problem as a principal-agent model but results are reconcilable with a bargaining model with incomplete information on the wife’s outside option.

4. Husbands observe whether their wives invest in the risky or the safe good and whether they use the purchased good (but not the good's productivity) and update their beliefs about their wives' expertise.

Choices In each period t of the spouses' lives, the husband chooses between making a transfer $T_t \in \{0, 1\}$ to his wife and enjoying the transfer ω himself. This transfer is in addition to a basic transfer determined by outside options as in the bargaining literature, which we normalize to zero. The wife then makes two binary choices: an investment choice, and an effort choice. Specifically, the wife decides whether to invest in a safe good, $g_t = 0$ (e.g., a well-known grain, medicine, or food, or possibly even just savings), or a new risky good, $g_t = 1$ (e.g., a new grain with potentially higher returns, a new medicine advertised to have fewer side effects, or a new food advertised to be more nutritious).⁵ The safe good has productivity $\eta_t = \eta^S$. The risky good has productivity $\eta_t = \eta_t^R$, where $\eta_t^R = \eta^R$ with probability $\lambda \in (\frac{1}{2}, 1)$ ⁶, $\eta_t^R = 0$ with probability $(1 - \lambda)$, and $0 < \eta^S < \eta^R$. The realization of the productivity of η_t^R is independent across periods t . Once the wife observes the productivity realization of the good she invested in, she decides whether to exert effort $e_t \in \{0, 1\}$ to use the good. The husband observes whether the wife invests in the safe or risky good (g_t) and whether the wife uses the purchased good (e_t), but he does not observe the return of the good ($y_t = e_t \eta_t$).⁷

The cost of using a good when its return is η_t is $c_t(\eta_t)$. For tractability, we set the usage cost to zero when the return of the good is non-zero: $c_t(\eta^S) = c_t(\eta^R) = 0$. The usage cost of the good is positive when the return of the good is zero: $c_t(0) = c > 0$. This is the "hiding cost" that women can choose to pay once they discover they purchased an unproductive risky good.⁸

Payoffs The goods are public goods in the sense that both spouses enjoy their payoffs. The payoff of the wife at time t is $v_t = T_t(y_t - c_t e_t)$. The (unobserved) payoff of the husband at time t is $u_t = T_t y_t + (1 - T_t)\omega$. Both spouses enjoy their payoffs at the end of their lives so that the final payoff of the wife is $V = \sum_{t=0}^{\infty} \beta^t v_t$ and that of the husband is $U = \sum_{t=0}^{\infty} \beta^t u_t$, where $\beta < 1$ is the discount factor.

We study the optimal behavior of both spouses under the assumption that their decisions are

⁵We assume that the types of goods for which wives receive transfers relate to specific purchases/types of investment for which women have a comparative advantage (due to, e.g., differences in information or differences in the opportunity cost of time). This implies that husbands never purchase such goods themselves.

⁶This ensures that increasing λ reduces the probability that non-expert wives make investment mistakes, as we show later.

⁷The assumption that the productivity of the good is not observable to husbands is not as extreme as it may seem: as will become clear later, goods whose low productivity can be observed over time are considered to have very high hiding costs, i.e., they are too costly to hide, so they will not be purchased in equilibrium.

⁸The model extends easily to a menu of usage costs $e_t \in \{0, e_L, e_H\}$, where $e_L > 0$ is required to use any good and $e_H > e_L$ is required to hide an unproductive risky good.

very frequent so that the discount factor β is very close to 1 (a reasonable assumption in our setting as wives go to the market several times per week):

Assumption 1. β is sufficiently high.

Wife Types As investors, wives can be either experts or non-experts, $\theta \in \{E, NE\}$. A wife's expertise is private information to the wife (let's say the wife learned her expertise through experimentation before marriage).⁹ Expert wives learn the realization of the risky good at the beginning of each period (before making their investment decision). Non-expert wives receive no additional information before investing: they only know that the risky good is productive with probability λ . It is only after investing in a good (but before they make their effort decision) that non-expert wives observe the exact realization of the good's productivity.

To illustrate the relevant mechanisms, we assume $\omega < \lambda\eta^R + (1 - \lambda)\eta^S$ and both $\omega > \lambda\eta^R$ and $\omega > \eta^S$. That is, the husband's outside option to making transfers is lower than the returns to always investing in the good with the highest return, but it is higher than the expected returns to always investing in the risky good or always investing in the safe good. Thus, if the husband perfectly observed his wife's expertise, he would make a transfer only if married to an expert wife.

The optimal action of the two types of wives without reputation concerns are the following: expert wives always invest in the risky good when it has high returns and invest in the safe one when the risky good is a lemon. Non-expert wives, who are not able to observe the productivity realization of the risky good before making their investment decision, invest in the risky good if and only if $\lambda\eta^R \geq \eta^S$.

2.1 Equilibrium Strategies

We next discuss the model's key intuitions. All results characterize Markov perfect equilibria, in which the husband decides whether to transfer money to his wife in each period t based on his belief about his wife being an expert, P_t . Let P_0 be the husband's prior (at the time of marriage) about his wife being an expert, equal to the (known) share of expert wives in the population. After each period, the husband uses Bayes' rule to update his beliefs about the wife's expertise based on her behavior. As the husband does not receive any further signals about his wife's expertise when he stops making transfers, we restrict ourselves to equilibria in cutoff strategies, implying the existence of a cutoff reputation level below which transfers stop forever.

⁹With many different types of goods, wives might be experts in some domains and non-experts in others, e.g., a wife could be an expert in farming investments but not in health investments. Thus, one should think of the model as applying to each domain separately, i.e., the husband decides independently whether to make transfers for farming purchases and whether to make transfers for health purchases.

We are interested in studying the dynamics and long-term equilibria when the realized return depends on usage, and the husband does not observe the productivity of the goods but observes his wife's usage of the goods. In this general case, the wife can "hide" a bad productivity draw by using a zero-productivity good. To build towards the general case, we first characterize the equilibrium strategies under two benchmark cases with no usage (the good is productive even when not used): 1) the husband does not observe the productivity of the risky good once the wife buys it, and 2) the husband observes the productivity of the risky good once the wife buys it. Once that is done, it is fairly straightforward to characterize the equilibrium strategies in the general case with usage.

2.1.1 Optimal Investment without Usage

First, we consider the case in which there is no usage. The realized return of the good is its productivity: $y_t = \eta_t$. Since the husband does not observe the return of the good and there are no effort decisions from which to learn, the husband receives no signal about the productivity of the risky good when purchased. In this case, the husband can only learn from the investment behavior of the wife.

Expert and non-expert wives have to decide strategically whether to invest, knowing that the husband cannot observe type I errors (the purchase of lemons) but only differences in investment rates over time. Let $\sigma_t^E(\eta_t^R, P_t)$ be the expert wife's investment strategy for the risky good (the probability the wife invests in the risky good) given the risky good's productivity realization η_t^R and the husband's belief about the wife being an expert, P_t ; let $\sigma_t^{NE}(P_t)$ be the non-expert wife's investment strategy for the risky good given her husband's belief about her being an expert, P_t .

We define the husband's value function and (Bayesian) updating functions, as well as the value functions of the two types of wives in theoretical appendix ???. Studying the optimal strategies of both types of wives in equilibrium, we derive the following lemma:

Lemma 1. *Suppose that the wife's prior reputation $P_0 \in (0, 1)$ is such that transfers occur ($\omega \leq P_0[\lambda\eta^R + (1 - \lambda)\eta^S] + (1 - P_0)[\sigma_t^{NE}(P_0)\lambda\eta^R + (1 - \sigma_t^{NE}(P_0))\eta^S]$) and that there is no choice over usage ($y_t = \eta_t$) s.t. the husband does not learn the productivity of the risky good. Then, in equilibrium, transfers occur every period and never stop (pooling). The average investment rates of expert and non-expert wives converge to λ as $t \rightarrow \infty$.*

The details of the proofs are in theoretical appendix ??. Here we present the key elements of the proofs.

The first step is to show that there exists a cutoff reputation. That is, the husband will make a transfer if and only if the reputation of the wife is higher than a certain threshold ($P_t \geq P^*$ for some P^*), below which the expected value of giving transfers to the wife is lower than the husband's outside option to giving transfers.

The second step is to show that if her reputation is above the cutoff reputation, the expert wife invests in the risky good when it is productive, that is at rate λ . Even if the expert wife deviated from this strategy when her reputation approaches the cutoff after a series of bad or good draws, her investment rate would still approach λ in the long-run.

The third step is to show that, to ensure that the reputation P_t does not fall below the threshold, non-expert wives' investment strategy converges towards the investment rate of expert wives as P_t tends to P^* when $t \rightarrow \infty$. In equilibrium, non-expert wives have a mixed-strategy that balances investing in the good with higher expected returns and investing in the good with higher expected reputation gains. P_t will fall quickly if the non-expert wife only invests in the good with expected higher returns; hence, this strategy is only possible when P_t is very far from the threshold P^* . The closer P_t gets to P^* , the more non-expert wives adapt their strategy to the investment rate of expert wives. Thus, to avoid being detected as non-expert, non-expert wives' investment strategy converges to λ in the long-run, the investment rate of expert wives. Thus, pooling occurs and the husband never stops giving transfers.

Lemma 1 shows the existence of a pooling equilibrium in which both types of wives manage to maintain their reputation above the cutoff reputation, and transfers never stop. This is possible because non-expert wives can easily conceal their type by mimicking the behavior of expert wives. Since the husband does not observe the productivity of the goods, he does not have any possibility of differentiating the two types of wives. In the next section, we discuss the case when the husband *does* observe the productivity of the goods.

2.1.2 Optimal Investment without Hiding

We now study the case where the return of the good depends on usage, $y_t = e_t \eta_t$, but the wife does not have the option to use low-return goods (no hiding). Thus, if the wife invests in the risky good, the husband immediately learns its productivity by observing whether it is used or not. This means that the husband is able to identify type I errors (the purchase of lemons).

In this environment, given that non-expert wives in expectation make more investment mistakes than expert wives, they cannot sustain the same investment rate in equilibrium as expert wives. Furthermore, if the husband has a near-zero prior belief about his wife being an expert, then he does not expect much valuable information from further trial to be generated as the posterior belief will be near zero as well by Bayes' rule. This drives the following lemma:

Lemma 2. *Suppose that the wife's prior reputation is $P_0 \in (0, 1)$ and that usage matters ($y_t = e_t \eta_t$) but wives cannot hide zero-return goods. Then, there exists a Markov equilibrium in which, as long as $P_t \geq P^*$, transfers occur. Investment occurs at different rates for the two types of wives with $\Pr(g_t = 1 | \theta = E) = \lambda > \sigma_t^{NE}(P_t \gg P^*) \geq 0$. Once $P_t < P^*$, transfers stop.*

The details of the proof are in theoretical appendix ???. The intuition of the proof is as follows.

The first step is to show that an equilibrium strategy in which non-expert wives invest more or equal than expert wives does not exist at $P_t > P^*$ and that non-expert wives avoid investing in the risky good unless they are close to the cutoff reputation. Since non-expert wives do not learn the productivity of the risky good before investing while expert wives do, non-expert wives incur the risk of making type I errors each time they invest in the risky good. This implies that an equilibrium in which non-expert wives invest more or as much as expert wives, and expert wives *never* invest in the risky good when it is not productive does not exist: non-expert wives would have an incentive to deviate since investing in the risky good implies a risk of losing transfers (in case they make a type I error) while investing in the safe good does not. If non-expert wives invest less in the risky good than expert wives, then the husband starts updating his posteriors based on his wife's investment behavior, and the wife's reputation decreases each time she invests in the safe good. Once her reputation approaches the cutoff reputation, the non-expert wife has two options: i) invest in the safe good and receive the state payoff now and no future transfers, or ii) invest in the risky good and potentially receive transfers in the future. With a sufficiently high discount factor, non-expert wives will prefer the second option.

The second step is to show that expert wives avoid type I errors as much as possible, meaning that any mistake will reveal the type of a wife. As long as the reputation of expert wives is not close to the reputation cutoff, expert wives invest in the risky good when it is productive and in the safe good when it is not. If they have a series of bad draws and their reputation approaches the cutoff, they have two options: i) invest in the safe good and receive the state payoff of the safe good but lose transfers in the future, or ii) invest in the risky good and commit a type I error. As non-expert wives will invest in the risky good when their reputation has reached the cutoff reputation, investing into the risky good is no longer a signal of being an expert wife. Furthermore, as non-expert wives are more likely to make type I errors than expert wives, making a type I error is a signal of being a non-expert wife. If expert wives choose option i), they will receive the payoff of the safe good. On the other hand, if expert wives choose option ii), their reputation will fall below the cutoff (as they will make a type I error, now a signal of being a non-expert wife) and they will lose transfers and receive a zero payoff. Thus, investing in the safe good when their reputation has reached the cutoff reputation maximizes their payoff. This means that even after facing a series of bad draws, expert wives will not make any mistakes and any type I error is thus a perfectly informative signal about the type of the wife.

The third step is to show that, given that expert wives make no type I errors, non-expert wives invest in the safe good as long as their reputation is above the reputation cutoff, at which point they start investing in the risky good. As expert wives make no type I errors, any type I error is a perfectly informative signal about the wife's type. Let's define the level of reputation at which the wife has

to stop investing in the safe good to continue receiving transfers ($P_t : T_{t+1}(P_{t+1}|g_t = 0) = 0$) in the next period as P^{**} . To avoid revealing their type and lose transfers forever, non-expert wives will thus invest in the safe good for as long as possible ($\sigma_t^{NE}(P_t \gg P^*) = 0$). At some point, they will have to increase their investment rate in the risky good to that of expert wives to maintain $P_t \geq P^*$ at the risk of losing transfers ($\sigma_t^{NE}(P^{**}) = \lambda$). The longer non-expert wives are able to maintain $P_t \geq P^*$, the closer $Pr(g_t = 1|\theta = NE)$ will be to $Pr(g_t = 1|\theta = E) = \lambda$.

2.1.3 Optimal Investment with Usage and Hiding

We now describe the general case in which the husband does not observe the return of the goods but only his wife's usage of the purchased goods.

Since the investment and usage (hiding) decisions take place sequentially, the stage optimal strategies are determined by backward induction. Wives have to decide strategically whether to invest and whether to exert effort. We define the husband's new Bayesian updating functions and value function, as well as the value functions for the two types of wives, in theoretical appendix ??.

Optimal effort depends on the hiding cost relative to the continuation value of receiving future transfers. When $\eta_t = \eta^R$ or $\eta_t = \eta^S$, the cost of usage is equal to zero and using the good is always optimal for both types of wife, as $\eta^R > \eta^S > 0$. When $\eta_t = 0$, the wife compares the expected benefit of hiding a bad purchase and keep receiving transfers to the hiding cost c . If at time t using a low return good is costlier than the gain the wife would get from using the good and keep receiving transfers in the future, given the equilibrium investment and usage strategies of both types of wives, then the same will be true for all future periods in which the good is non-productive (because the continuation value is decreasing over time). This implies that hiding never happens. Without hiding, the husband learns the productivity of the good and the proof of lemma 2 applies. When, instead, the cost of hiding is low and the future expected pay-off is higher than the cost of hiding, the wife will use the low-return good to keep receiving transfers. Thus, the husband never learns the productivity of the good and the proof of lemma 1 applies. Thus, we have the following proposition:

Proposition 1. *When the hiding cost is low, the investment rate of the two types of wives converges as $t \rightarrow \infty$ (pooling); non-expert wives hide whenever they commit a type I error. When the hiding cost is high, non-expert wives never hide through usage and there exists a Markov equilibrium in which, as long as $P_t \geq P^*$, transfers occur. Investment occurs at different rates for the two types of wives (separation), with non-expert wives investing less often than expert wives. Once $P_t < P^*$, transfers stop.*

2.2 Comparative Statics of the Equilibrium

We have shown in the previous subsection that in the setting with hiding, the hiding costs determine whether non-expert wives are able to receive transfers indefinitely (pooling) or will stop receiving transfers at some point (separation). Before we turn to the testable predictions we take to the data, we show how the equilibrium investment strategies vary with λ , the probability that the risky good is productive, and ω , the husband's outside option. The derivations are shown in theoretical appendices ?? and ??. The formal corollaries are as follows:

Corollary 1. *When risk decreases (λ increases), investment of both types of wives in the risky good increases and, for high hiding costs, the difference in investment rates between expert and non-expert wives decreases as $t \rightarrow \infty$. If $\lambda = 1$, the investment of both types of wives in the risky good is 1.*

Corollary 2. *The cutoff reputation below which transfers stop, P^* , increases in the husband's outside option to transfers ω . This implies that, for high hiding costs, non-expert wives start investing in the risky good earlier.*

2.3 Testable predictions

In the previous subsections, we formally showed how in the setting in which husbands only observe usage and hiding is possible, wives' optimal investment strategies vary with the cost of hiding investment mistakes. These theoretical results give us predictions that we empirically test through a series of experiments in Malawi. Since our experiments were run with real couples/wives, we consider our activities as being one particular round of a dynamic game with infinite horizon. We test the following key premise and predictions:

Premise 1. *Husbands' financial transfers to their wives respond to their beliefs about their wives' expertise as investors.*

Prediction 1. *When the price of hiding is low, non-expert wives invest no less than expert wives (pooling equilibrium in investing). When the price of hiding is high, non-expert wives invest less than expert wives (separating equilibrium in investing).*

Prediction 2. *Non-expert wives, conditional on investing, hide investment mistakes (over-use of lemons).*

Prediction 3. *The difference in investment rates between experts and non-expert wives decreases as uncertainty about the quality of the risky good decreases.*

Prediction 4. *Spouses' behavior stops responding to the wife's reputation once it has fallen below the cutoff reputation above which discretionary transfers occur.*

The rest of the paper tests these predictions. First, in section 4, we test premise 1 in a lab-in-the-field experiment in which husbands transfer money to their wives and we randomly vary the salience of the wife's market expertise reputation. Second, in section 5, we use a complementary lab-in-the-field experiment to test predictions 1 and 2. We do so by randomly varying hiding costs and studying investment and hiding for wives. Third, in section 6, we test predictions 1 and 3 in a field experiment involving a real-life purchase decision: we offer an unfamiliar good to women while they are running errands at the market, randomly varying the hiding cost and the observed uncertainty about the good. Finally, as predictions 1 to 3 only apply to couples for which discretionary transfers still occur, we test prediction 4 by assessing heterogeneity in the size of the transfer from husbands to wives in all three experiments. This also allows us to verify that behaviors in the experiments are externally relevant: driven by concerns over real-life transfers, and not merely experimenter demand effect.

3 Empirical Setting: Couples in Rural Malawi

The transfer experiment and the signaling experiment were done side-by-side with 1,093 married couples between May and July 2019. The couples were sampled from 36 villages in Neno district, in Southern Malawi. We selected dwellings randomly and enrolled households in which both spouses were available to participate in an hour-long survey administered separately to the husband and the wife.¹⁰ The husband survey embedded the *transfer experiment*. The wife survey embedded the *signaling experiment*. Everything took place at the couples' homes, with one surveyor speaking with the husband while the other spoke with the wife. Interviews were held outdoors and far enough apart to respect complete confidentiality for both spouses.

Table 1 provides some summary statistics on the couples surveyed. The surveys administered included standard questions on household demographics, schooling, and employment, as well as a module on expenditures and budget decisions inside the household, recent transfers from the husband to the wife, and financial literacy. In addition, we elicited respondents' performance on six math questions to test respondents' ability to solve everyday math problems, and on 12 Raven's Progressive Matrices (Cattell, 1963) to measure respondents' reasoning ability.

The couples in our sample have been married for an average of 10 years and have 2.6 children. Husbands are an average of 36 years old, have 6.8 years of education and have earned an average of Malawian Kwacha (MWK) 29,769 (approx. USD 44) per month in the preceding two months. Wives are an average of 30 years old, have 5.7 years of education and have earned an average

¹⁰Enumerators used the "left-hand" rule to randomly sample dwellings, as described in appendix B.1.

of MWK 10,643 (approx. USD 16) per month in the preceding two months. Husbands report transferring an average of MWK 8,428 (approx. USD 13, 30% of their income) per month to their wives in the preceding two months.

Table 1: Couple characteristics, transfer and signaling experiments

Variable	Obs	Mean	Std. Dev.
Years married	1092	9.91	8.48
Number of children	1093	2.63	1.56
Husband's age	1093	35.83	10.16
Husband's education	1093	6.77	3.54
Husband's avg. income last two months (MWK)	1093	29769.24	33071.78
Wife's age	1091	30.37	8.93
Wife's education	1093	5.68	3.27
Wife's avg. income last two months (MWK)	1093	10643.17	17476.74
Avg. transfers (H to W) last two months (MWK)	1093	8427.66	11306.75

Notes: Kwacha values are winsorized at 3 SDs. They represent averages over the preceding two months.

The *market experiment* was done with a separate sample a few months later (July 2019). We discuss the sampling strategy of that experiment in detail in section 6, but the two samples have very comparable characteristics.

4 Does Reputation Matter for Budget Shares? The Transfer Experiment

4.1 Measuring Market Expertise Reputation

During the survey with the husband, we elicited his beliefs about his wife's market expertise. From this, we construct the wife's Market Expertise Reputation (MER) index, that takes the values 0 to 4, depending on how many of the following questions the husband affirmed: i) his wife has never bought anything that did not work as advertised ("Purchases", 86%), ii) his wife is never tempted to buy advertised goods with uncertain return at the market ("Tempted", 80%), iii) he believes his wife can manage money well compared to other women in the community ("Manage", 70%), and iv) his wife can do calculations correctly in her head when she requests change in the market ("Change", 95%). The distribution of the MER index is as follows: 0.6% of women have an MER of 0, 3.5% have an MER of 1, 13% have an MER of 2, 31% have an MER of 3, and 52% have an MER of 4.

When asked to recall their wife's purchase behavior, husbands were asked to provide examples of instances when the wife was tempted to make a purchase of an advertised good and instances

when a good purchased by the wife did not work as advertised. Husbands were easily able to recall such instances. Online appendix table C.1 lists 50 randomly selected answer choices to each of those questions. The main “flaw” of non-expert wives appears to be gullibility, in the sense that they get easily fooled by vendors. Examples of direct quotes from the husband surveys include: “She bought a drug that wasn’t effective at all. She got carried away by what the vendor was telling her”; “She bought a pair of shoes that were not of the required foot size because a vendor told her it will fit”; “She bought actellic ‘super dust’ that didn’t work”; “She bought a dish, which was looking nice but it broke down in just a month”; “She bought second-hand burglar bars, which were painted to conceal the rust”; “She was given short trousers by the vendor instead of a skirt.”

Our model predicts that the amount of money a wife receives from her husband is positively correlated with her intra-household reputation as an expert. This appears to be the case observationally, as shown in appendix table 6: controlling for husband and wife characteristics, reported average transfers in the previous two months are monotonically increasing in the wife’s MER. On average, women with an MER of 0 or 1 (4%) receive MWK 8,902, women with an MER of 2 (13%) MWK 9,261, women with an MER of 3 (31%) MWK 10,143, and women with an an MER of 4 (52%) MWK 10,692 (column 1). To verify that men transfer discretionary funds to their wives according to a cutoff strategy, we define women as “High MER” if they have an MER of 3 or 4 (83%). Indeed, having a high MER of 3 or 4 is associated with an increase in reported average transfers in the previous two months by MWK 1,261 (15%, column 2). These results are robust to controlling for an indicator that is 1 if the wife has an above-median “General Ability Reputation” (GAR), a mean effects index (see [Kling et al. \(2007\)](#)) of the husband’s beliefs of the wife’s scores on six math questions and 12 raven matrices (columns 3-4). In column 5, we look at another measure of how much control the wife has over the budget: whether the husband reports that his wife has access to cash and savings. Here again we find a significant, positive and large correlation with high Market Expertise Reputation.¹¹

The fact that most husbands believe their wives are experts suggest that either the share of experts in the population is very high, or that there is a good deal of “pooling” as predicted by the model. Nevertheless, the MER as reported by husbands seems correlated with true types: 73% of wives with a low MER report that the statement “I buy things that I later regret because I bought them on impulse” applies to them. This share falls to 62% among wives with a high MER (the p-value of the difference is <0.01).¹²

¹¹As we show in the next subsection, our results in the *transfer experiment* are also consistent with an equilibrium in cutoff strategies with a cutoff MER of 3.

¹²The correlation between high MER and risk-aversion of the wife is close to 0, suggesting that high MER does not simply proxy risk aversion. The results are robust to controlling for a measure of risk aversion: the response to the question “You are buying 10 packets of seed. There are two kind of seeds. One will mature with a low yield, but it is guaranteed that all seeds will mature. The other will mature with a higher yield: You will get 3x as much. However, on average, half of all packets of risky seeds never mature. That means you could have all good packets or all bad packets

4.2 The Transfer Experiment - Experimental Design

The observational results above are suggestive, but do not nail causation. The correlation between reputation and transfers could be driven by, e.g., recall bias: men who hold their wives in high esteem could be more likely to remember a transfer to their wife. For a clean test of premise 1, we implemented a “transfer experiment” in which we asked husbands to play a dictator game with multiplier (with their wife as the counterpart), and experimentally varied the salience of the wife’s reputation at the time the husband made his choice.

Specifically, in the transfer experiment, husbands were offered to choose what share of their experiment compensation would be doubled and transferred to their wives.¹³ Husbands were randomly assigned to either of two versions of the experiment:

- *Control*: Husbands played the dictator game early in the survey, before beliefs about their wife’s market expertise had been solicited.
- *Salience Treatment*: Husbands played the game at the end of the survey, immediately following the MER module asking them to recall their wife’s purchase behavior.

Husbands were explained the dictator game as follows:

“As promised, we are going to give you 600 mk for participating in the survey. Here is the 600, please count it to make sure it’s correct. But before you pocket it, I am going to offer you a chance to give some of this 600 mk to your wife. Here is how it will work. You will choose how much of the 600 mk you want to give to your wife. Whatever you choose to give her, we will double. So if you give 20 mk to your wife, we will give her 40 mk and you keep 580 mk. If you give 400 mk, we will give your wife 800 mk and you keep 200 mk. [...] If you choose to give money to your wife, she will get it right away. We will not tell her where this money is coming from: We will only tell her that this is part of the survey. If you choose to give 0, we will not tell her anything at all.”

To reduce the risk of experimenter demand effects, respondents placed the money they wanted to donate to the wife in an envelope, which was then handed (with the husband bearing witness) to the surveyor speaking with the wife (without the wife bearing witness).

While our model primitives predict that husbands should transfer less to women with a non-expert reputation in a standard dictator game (as we have already shown they do observationally), we multiply all transfers to the wife to give all husbands an incentive to transfer to their wives. By randomizing how “top of mind” the wife’s (potential lack of) market expertise is at the time husbands make their transfer choice, we can then obtain a causal estimate of the importance of reputation in the husband’s allocation decision.

To test this, we estimate the following equation:

or somewhere in between. Out of the 10 packets you will buy, how many packets of the risky seed will you buy?”

¹³See technical online appendix B.2 about additional information about the compensation.

$$T_i = \alpha + \beta_1 \text{LowMER}_i + \beta_2 S_i + \beta_3 (\text{LowMER}_i \times S_i) + \beta_4' X_i + \mu_e + \delta_c + \lambda_v + \epsilon_i \quad (1)$$

where T_i is the dictator game transfer of husband i to his wife, LowMER_i is a binary variable equal to 1 if the wife’s Market Expertise Reputation is low, and S_i is assignment of husband i to the salience treatment. ϵ_i are Huber-White robust standard errors. We include enumerator fixed-effects μ_e , compensation fixed effects δ_c and version fixed effects λ_v (see technical online appendix B.2 for details on what the different compensations and versions are). We show results both without and with a vector of predictive individual controls X_i .

Since we elicited the MER at the same time as we implemented the experiment, we could not stratify the salience treatment by MER level. Nevertheless, within each MER group, treatment and control groups appear balanced (online appendix table C.2).

4.3 The Transfer Experiment: Results

Table 2 presents the results of the transfer experiment. Husbands transfer 69% of the funds to their wives on average—a relatively large share compared to the outcomes of similar games in most other contexts, which are usually not played within couples (Andreoni and Vesterlund, 2001; Andreoni and Miller, 2002; Jakiela, 2013). Making the wife’s lack of market expertise salient just before the transfer reduces transfers: the salience treatment decreases the transfer share by 9.1ppts (13%, $p < 0.05$, column 2) among women with a low MER, but it does not change anything for women with a high MER. In other words, the salience module made husbands of low-MER wives think twice and adjust their transfer downward from what they would have done otherwise.¹⁴

The experimental results demonstrate that transfers to the wife respond substantially to increasing the salience of the wife’s perceived reputation. Taken together with the fact that real-life transfers are also correlated with the wife’s perceived reputation, this is consistent with premise 1 of the model: husband transfers seem to respond to husbands’ beliefs about their wives’ expertise as investors.

¹⁴We look at the individual components of the high MER index one by one as well as at the different MER subsamples in appendix table C.3. The effects are driven by wife’s past experience with bad purchases and temptations. Consistent with the equilibrium result that husbands adopt a cutoff strategy, the salience treatment decreases transfers among women with a score of 2 or below but it does not decrease them among women with a score of 3 or 4. We also conduct robustness analysis in table 7. We rule out experimenter demand effects by looking at whether the salience treatment influences transfers based on what husbands said about other aspects of their wives’ ability, notably their math and raven’s score (grouped in the GAR index). Another concern could be that husbands who just had to recall mistakes made by their wives were *angry* at the time they decided on the transfer, and anger is what caused them to reduce contributions, not reputation of the wife *per se*. We check whether husbands who may have gotten angry due to another section of the survey (those who scored poorly on the math and raven’s quiz) transfer less and find it is not the case. Finally, as we show in section 7, consistent with the theoretical prediction, husbands who stopped giving discretionary transfers to their wives do not react to the salience treatment.

Table 2: Transfer experiment: Effect of reputation salience on amount (%) transferred from the husband to the wife

	(1)	(2)
Salience	2.143 (1.720)	1.723 (1.736)
Low MER	0.343 (3.374)	-0.463 (3.376)
Low MER*Salience	-9.184 (4.231)	-9.084 (4.176)
Mean (Control & Low MER)	68.889	68.889
Observations	1093	1093
Controls		✓

Notes: The table shows results from OLS regressions with Huber-White robust SEs. Market Expertise Reputation (MER) is an index that takes the values 0 to 4 (see main text). All regressions include enumerator, compensation and version fixed effects. Controls include wife and husband’s age, education, average income and transfers in the last two months, variability of income (whether income is the same in most months or varies a lot), risk preferences, math and raven scores, and years married, number of children and number of household members.

5 Are Women Strategic? The Signaling Experiment

Having shown that husbands act on their beliefs about their wife’s expertise as a buyer, we now turn to testing prediction 1—non-expert wives invest as much as expert wives when the price of hiding is low (*pooling*) but less when the price of hiding is high (*separation*), and prediction 2—non-expert wives, conditional on investing, hide more than expert wives. We do this through a signaling experiment implemented with the wives of the men who participated in the transfer experiment.

5.1 Experimental Design

At the end of a survey for which they were paid a compensation,¹⁵ women were asked to complete a “quality quiz”. The quiz was designed to mimic a situation in which a person has to assess the quality/productivity of a good in order to decide whether to purchase it. In each of six

¹⁵See technical online appendix B.2 for more information about the compensation.

rounds (and two practice rounds) respondents were shown a high-quality and a low-quality version of a good, and asked to guess which of the two versions was higher quality. The six goods were a sponge, a water bottle, a razor, a toothbrush, flour and a candle, and the order in which the goods were presented was randomized. On average, both women and men identified 4.2 high quality goods correctly.

After they had completed the quiz, we elicited respondents' prior distributions about their performance on the quiz, using beans to represent probabilities and visual aids to represent the support (Delavande et al., 2011). We use the mean prior of this elicitation to classify women as follows:

- “*Self-Identified Expert*”: Women with a mean prior (averaged across the 7 possible scores) of 5 or 6 (61% of the sample). These are women at lower risk of sending a “bad” signal to their husbands if they choose to participate without hiding.¹⁶
- “*Self-Identified Non-Expert*”: Women with a mean prior (averaged across the 7 possible scores) of below 5 (39% of the sample). These are women at greater risk of sending a “bad” signal to their husbands if they choose to participate without hiding.¹⁷

After taking the quiz, but before seeing her initial score, the wife was given the option to participate in an extra activity for an additional compensation (worth 50% of the compensation already promised for answering the survey). This extra activity was as follows. If the wife agreed to participate in the activity, we would inform her husband (at the end of the survey) about her score on the quality quiz, but first she would be given the opportunity to improve her score by choosing the number of potential mistakes she would want to correct against a fee. In the terms of the model, we gave wives a costly opportunity to hide some of their mistakes. For the signal to be meaningful, husbands were administered the same quiz as part of the survey they did during the transfer experiment. Women were told so, and also told that their husband had been asked his prior about the wife's score on the quiz.¹⁸ Husbands, however, were not told that the wife had a *choice* over a signal possibly being sent. If the wife chose to participate, the husband received the wife's final (post-hiding) score on the quiz (not knowing she had the option to pay to improve her score first). If the wife chose not to participate, the husband did not receive any information at all about the wife nor about the content of the wife's survey.

We elicited wives' decisions in the following order: First, we provided the wife with the unit price she would have to pay to correct a mistake in the quality quiz. We clearly spelled out how many answers she would be able to afford using the experimental payments (the participation fee for the

¹⁶This is smaller than the share of wives *perceived* as experts by their husband (wives with high MER) in section 4, further suggesting that pooling may be at play.

¹⁷Results are almost identical and statistically indistinguishable when using the modal, minimum or maximum prior.

¹⁸Men and women performed equally well on the quality quiz. See detailed scores in appendix table C.4.

extra activity itself + the participation fee for the survey). Women could not purchase answers with their own funds and we made that very clear before they had to make their decision to participate. Second, we asked the wife to decide if she wanted to participate in the activity (i.e., whether she would let us tell her final (post-hiding) score to her husband), given her prior about her score and the correction (hiding) fee schedule. Third, among those women who were willing to participate for money, we elicited how many mistakes they would want to correct under each possible scenario (i.e., in case they had 0, 1, 2, ... questions correct). Since the correction fee was deducted from the compensation fee yet to be paid out, the elicitation was incentive-compatible (women could not renege on their correction choice after seeing their score).

Respondents were assigned to one of the following hiding price treatment arms:

- *Low hiding price*: The low price allowed wives to correct no more than two mistakes when using the additional compensation from participating in the activity and up to six (i.e., all possible) mistakes when using the additional compensation from the activity as well as their survey compensation.
- *High hiding price*: The high price allowed wives to correct no more than one mistake when using the additional compensation from participating in the activity and no more than three mistakes when using their entire compensation.

The experiment was designed to test predictions 1 and 2 of the signaling model. Specifically, the only rational reason *not* to participate in the extra activity despite the compensation offered is to avoid sending a bad signal to the husband. Likewise, the only rational reason for giving up part of the compensation to hide mistakes is to avoid sending a bad signal to the husband. By observing non-zero rates of non-participation and hiding, we can already assert that women are concerned with their reputation. By varying the hiding fee randomly, we are able to test specific predictions from the model. Specifically, whether

1. Prediction 1: non-expert wives invest no less than expert wives when the price of hiding is low (pooling equilibrium in investing) but less when the price of hiding is high (separating equilibrium in investing).
2. Prediction 2: non-expert wives, conditional on investing, hide more than expert wives (over-use of lemons).

“Investing”, in the experimental context, consists of accepting to participate in the extra activity where a signal is sent to the husband. “Hiding” consists of paying to correct mistakes before the signal is sent.

We estimate the impact of the high hiding price on the participation and hiding behavior by self-identified expertise status using the following equation:

$$Y_i = \alpha + \beta_1 NE_i + \beta_2 HP_i + \beta_3 (NE_i \times HP_i) + \beta_4' X_i + \mu_e + \delta_c + \epsilon_i \quad (2)$$

where Y_i is outcome for woman i . NE_i is as an indicator that is 1 if the wife self-identifies as non-expert (her average prior about her performance is below 5) and HP_i is an indicator that is 1 if woman i is assigned to the high hiding price. ϵ_i are Huber-White robust standard errors. As above, we include enumerator fixed-effects μ_e , and compensation fixed effects δ_c .¹⁹

To address the potential concern that results in the experiment are driven by the endogeneity of the wife's self-identified expertise, we created exogenous variation in performance on the quiz by further randomizing wives into two groups as follows:

- *Hard version*: Respondents did the quality quiz without hints.
- *Easy version*: Respondents were provided hints to help them discern the high-quality good during the quiz.

Online appendix table C.4 documents that the hints succeeded in increasing women's performance in the quiz: they significantly increased the share of women who correctly discerned the high-quality good for each of the six goods, thereby increasing the average score by 1.1 points from 3.6 points (out of 6) in the hard version to 4.7 points in the easy version. In addition, the hints also significantly increased respondents' priors about their performance. To verify that this also reduced wives' reputation concerns, we elicited their second order beliefs (beliefs about their husbands' beliefs about the wife's performance in the quiz), using the visual handouts shown in online appendix figure C.1. As intended, the hints substantially reduced the share of wives who believed that their husband would update negatively about his wife's expertise if she participated in the signaling activity (reported a mean prior below their 2nd order beliefs): 22% of wives playing the easy version vs. 31% of wives playing the hard version.

This randomized variation in performance allows us to test whether wives with an exogenously lower performance participate less in the signaling activity when hiding is costly, and pay to hide their mistakes when hiding is cheap. We do so by estimating the following equation (similar to equation 2):

$$Y_i = \alpha + \beta_1 Hard_i + \beta_2 HP_i + \beta_3 (Hard_i \times HP_i) + \beta_4' X_i + \mu_e + \delta_c + \lambda_s + \epsilon_i \quad (3)$$

where $Hard_i$ is as an indicator that is 1 if the wife was assigned the hard quiz version and HP_i is the indicator that is 1 if woman i is assigned to the high hiding price. ϵ_i are Huber-White robust standard errors.

¹⁹Controlling for whether the husband was assigned to the salience treatment in the transfer experiment described in section 4 does not change the results, as expected since the flow of the wife survey and embedded experiment were completely independent of that of the husband's.

5.2 Results

Table 3, panel A, presents the results exploiting the price variation only, with the full sample. Results with controls are shown in online appendix table C.5, panel A. Column 1 shows that self-identified non-expert wives have a significantly lower score than self-identified expert wives (3.9 vs. 4.1, $p < 0.10$). Note that this is not mechanical—women could be off about their own expertise, but the data suggests they are not. The next two columns show that a substantial share of women forgo experimental earnings in order to avoid sending a signal. To start, only around 76% of women decide to participate even if the price of hiding is low (column 2).²⁰ Non-expert wives participate as much as expert wives if the price of hiding is low. However, the high hiding price decreases the share of non-expert wives that participate by 11.9 percentage points (-16%, $p < 0.05$, column 2). This is completely in line with our theoretical prediction 1: participation rates do not differ by self-identified expertise when the price of hiding is low (pooling equilibrium) but diverge when the price of hiding is high (separating equilibrium). Wives with a low perceived score forgo 71 MWK in participation fee on average (column 3).

Conditional on participating, non-expert wives still have a significantly lower score than expert wives (-0.36 points, $p < 0.05$, column 4). As expected from theoretical prediction 2, non-expert wives who participate hide significantly more than expert wives who participate: they correct 0.25 more errors ($p < 0.05$, column 5), thus paying substantially more in hiding fees (column 6). The high price of hiding reduces hiding among both non-expert and expert wives, and only partially reduces the hiding gap between them. Ultimately, given the differential hiding rate at the low price, final (corrected) scores (sent to husbands) are statistically indistinguishable between non-expert and expert wives, hence the husband receives an uninformative signal about his wife’s expertise (column 7).

Wives forgo experimental earnings on both the extensive (participation) and intensive (hiding) margin in order to avoid sending what they think will be a bad signal. Self-identified non-experts forgo 30% more earnings (MWK +24.5, from a base of 82.6) than experts when the price of hiding is low, and 41% (MWK +36.9, from a base of 90.7) more when the price of hiding is high (column 8).²¹

To address the concern that results could be driven by the endogeneity of type (for example, non-experts forgo more because they are worse at math), we present results with exogenous variation in performance in table 3, panel B. The harder version of the quiz succeeded in exogenously lowering women’s scores from 4.6 points to 3.5 points (column 1). When the price of

²⁰Online appendix figure C.2 documents that participation rates are increasing substantially in the women’s reported risk preferences. All our results are robust to controlling for risk preferences (see online appendix table C.5).

²¹Since our predictions and key estimations concern the interaction between expertise status and the randomly assigned hiding price, the relevant “balance check” is within each expertise group. We show this in online appendix table C.6.

hiding is high, wives with a randomly lowered performance are less likely to choose to participate in the signaling experiment (-9.0 pts, column 2) in exchange for the participation money (column 3). Conditional on participating, they hide an average of 0.3 additional mistakes ($p < 0.01$, column 5) and pay an average additional hiding cost of MWK 31 ($p < 0.01$, column 6). This is not enough to close the gap in scores (column 7).

Overall, wives who were randomly thrown a hard ball incurred an increase in forgone earnings of 14% ($p > 0.10$, column 8) when the price of hiding was low and of 45% ($p < 0.10$, column 8) when the price of hiding was high. This is consistent with wives in the hard quiz treatment assuming that their husbands will interpret their scores on an absolute scale (not adjusting for the difficulty of the quiz), and wanting to avoid sending a bad signal.

Through the lens of the model we propose, wives in the signaling experiment are willing to forgo experimental earnings to protect their reputation in order to maintain access to financial transfers. An alternative is that they need to protect their reputation in order to avoid other negative consequences from a low reputation, such as domestic violence or emotional abuse. The model encompasses such alternative: it can be thought of as a version of the model where transfers are not only financial but also in-kind (with abuse being a negative transfer). We therefore see the results from the signaling experiment as supportive of the model in any case. From an ethical standpoint, however, the implications of our experimental design are quite different across the two interpretations. Specifically, if the risk of abuse increases in response to poor investment choices, could our signaling experiment have put our participants at risk?²² We piloted the protocols extensively in settings where women could freely share with us their concerns, and abuse was never brought up. Our enumerators, after sharing the wife's performance with the husband (with the wife present), systematically witnessed husbands congratulating their wives on their good performance (the wife's final reported score was 4.6 out of 6 on average, higher than the husbands' average of 4.2). Furthermore, the results of our heterogeneity analyses are much better aligned with the hypothesis that women attempt to maintain their reputation to maintain financial transfers, than with a fear of violence mechanism. First, we find strong heterogeneity by transfers, suggesting that women who could lose more financially also respond more to our experiments (see section 1). Second, the 2015/2016 Malawi DHS shows that violence is more likely to occur later on in a couple's marriage (table 17.13, [NSO and ICF \(2017\)](#)) but we find that our results are driven by women early in their marriage.²³

²²Domestic abuse is present but not widespread in our study setting: The share of adult women in Neno District who reported experiencing physical or sexual intimate partner violence, in the past 12 months is 12% ([NSO and ICF, 2017](#)), compared to 25% in Kenya ([Kenya National Bureau of Statistics and Ministry of Health and ICF, 2017](#)), 27% in Bangladesh ([Bangladesh Bureau of Statistics and Bangladesh Bureau of Statistics, 2016](#)), 9% in Guatemala ([Ministerio de Salud Pública y Asistencia Social and Instituto Nacional de Estadística, INE](#)) and 5% in the US ([Smith et al., 2018](#)).

²³The coefficient of "Harder Version*High Price" on total foregone income is 58.62 for younger couples (< 7 years) and 7.33 for older couples (≥ 7 years).

Table 3: Outcomes in the signaling experiment

	Whole sample (N=1093)			Participation sample (N=786)				Whole sample (N=1093)
Panel A: By price and low perceived score								
	Initial score	Participate (%)	Foregone comp.	Initial score	Errors corrected	Hiding fee	Final score	Total forgone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Self-Identified Non-Expert	-0.246 (0.129)	-4.082 (4.666)	8.165 (9.331)	-0.361 (0.147)	0.247 (0.108)	23.283 (11.021)	-0.114 (0.158)	24.535 (10.763)
High Price	0.138 (0.097)	1.948 (3.355)	-3.897 (6.710)	0.146 (0.110)	-0.209 (0.065)	14.171 (9.329)	-0.063 (0.113)	8.136 (8.960)
Self-Identified Non-Expert*High Price	0.065 (0.156)	-11.874 (5.772)	23.749 (11.544)	0.091 (0.180)	-0.146 (0.121)	1.121 (17.240)	-0.055 (0.190)	12.418 (14.519)
Mean (Low Price & Self-Identified Expert)	4.149	76.349	47.303	4.196	0.462	46.196	4.658	82.573
P-value (Self-Identified Non-Expert vs. Expert, High Price)	0.041	0.000	0.000	0.009	0.073	0.068	0.106	0.000
Panel B: By price and difficulty of the quiz								
	Initial score	Participate (%)	Forgone comp.	Initial score	Errors corrected	Hiding fee	Final score	Total forgone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Harder Version	-1.120 (0.111)	6.859 (4.428)	-13.718 (8.856)	-1.064 (0.129)	0.306 (0.100)	30.765 (10.242)	-0.758 (0.143)	11.561 (10.369)
High Price	0.108 (0.090)	1.208 (3.944)	-2.417 (7.888)	0.130 (0.105)	-0.205 (0.064)	5.375 (9.216)	-0.075 (0.109)	0.903 (9.306)
Harder Version*High Price	0.121 (0.137)	-8.976 (5.512)	17.951 (11.025)	0.067 (0.157)	-0.098 (0.112)	19.665 (15.956)	-0.031 (0.172)	27.104 (14.225)
Mean (Low Price & Easy Version)	4.594	71.875	56.250	4.601	0.406	40.580	5.007	85.417
P-value (Hard vs. Easy Version, High Price)	0.000	0.522	0.522	0.000	0.000	0.000	0.000	0.000

Notes: The table shows results from OLS regressions with Huber-White robust SEs. Low Perceived Score is an indicator that takes the value 1 if the wife reports an average weighted score that is lower than 5 (39 % of women). The weighted average is calculated as the average across all scores, weighted by the probability assigned to each score by the woman (each woman placed 10 beans on the 7 different scores). Foregone comp. is the amount of money wives left on the table by opting out of the game. All regressions include enumerator and compensation fixed effects. The p-value is the p-value from a Wald test comparing outcomes between low perceived score and high perceived score wives or between the hard and the easy version when the price of hiding is high.

6 The Market Experiment

Following the completion of our lab-in-the-field experiments, we conducted short surveys and a field experiment with 675 married women, recruited while they were shopping at one of six local markets in Zomba district in July 2019. This experiment tests prediction 1 for real-life investment rates—do non-experts invest no less than experts when the price of hiding is low but less when the price of hiding is high?, and prediction 3—does the difference in investing probabilities between experts and non-experts decrease as the uncertainty about the return of the risky good decreases?

We recruited participants on market days as follows. We approached women who were at the market to make purchases, and were alone. We first administered a short survey that included standard questions on household demographics, schooling, and employment, as well as modules on expenditures and budget decisions in the household, and recent transfers from the husband to the wife. The women included in the market experiment have been married for an average of 10 years and have 2.5 children (see online appendix table C.7). They are an average of 30 years old, and have 7.3 years of education. They have earned an average income of MWK 15,117 (approx. USD 23) in the preceding two months, are married to husbands who have earned an average income of MWK 20,397 (approx. USD 30) in the preceding two months, and have received an average transfer from their husband of MWK 11,293 (approx. USD 17) in the preceding two months.

6.1 Experimental Design

We offer women a compensation for their survey time. Using a Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964), we elicit their willingness to trade part or all the cash value of the compensation for an unfamiliar good. This generates an estimate of their willingness to pay for the unfamiliar good, without the amount of cash on hand they have at the time of the survey being a constraint.

Women were randomized into being offered one of two unfamiliar goods:

- *Low hiding (time) cost*: An airtight crop storage bag purchased in Blantyre, the second largest city in Malawi. These bags are hermetically sealed to protect harvested grains (e.g., maize, red beans) from insect pests. They are used to store grains for months on end. While the returns of these bags are substantial, they were unknown to our respondents. The usage cost of a storage bag is low (in terms of the model, the cost of hiding a wrong investment is low). This is because once the bag has been filled, even if it did not truly protect from pests, leaving the grains in the bag for the rest of the season would have no cost (since the status quo is to store the grains in non-sealed bags).
- *High hiding (time) cost*: A child picture book (imported from overseas by the research team):

Either Richard Scarry’s “A Day at the Airport” picture book, or the illustrated “Lift-the-flap” Animal ABC baby book by Jonny Lambert. Such books are relatively expensive (USD 10 before shipping costs) and totally unavailable in Malawi, even in the capital city. Even low quality picture books are completely absent from local markets. The evidence on the benefits of showing books and describing pictures to very young children is now strong; however, parents worldwide typically underestimate the returns of this kind of activities. Board books are meant to be looked at/shown/read to children over and over again. The usage cost (cost of hiding) is thus high, because the good needs to be used repeatedly.²⁴

We implemented the intervention in market spots so that, in case the respondent got the good and brought it home, the husband would infer that the respondent had bought the good at the market. To create further exogenous variation that helps test the model, women were randomly allocated to one of four sticker groups:

- *Donated*: We put a sticker on the good (e.g., back-cover of the books) stating “donated by Stanford University”.
- *Effectiveness*: We put a sticker on the good describing the proven effectiveness of the good (e.g., explanations on the positive effects of reading/looking at picture books with children).
- *Both*: We put both Donated and Effectiveness stickers.
- *None*: No stickers.

The stickers are shown in online appendix C.1.1.²⁵ When deciding whether and how much to invest, the women could see the stickers. Hence she knew what information would be available to her husband. Specifically:

- The donated sticker gave the woman the guarantee that the spouse would see that the good was acquired at no financial cost (as it indeed was, since it was given in exchange for her time). In such a case, the reputation mechanism in the model is not at play as the cost of the good to the husband is 0, i.e., the payoff to the husband cannot be negative. The prediction is that investment rates should differ across experts and non-experts absent the sticker, but not in the presence of the sticker.
- The effectiveness sticker aimed to inform the husband about the productivity of the good (lemma 2) and reduce uncertainty about the quality of the risky good to the husband (corollary

²⁴Because we had to bring the books by plane in a suitcase (there is no reliable shipping service to Malawi), we were only able to offer the books to 26% of the women in the sample.

²⁵Even though literacy is high in our setting (88% among adult men and 78% among adult women in Zomba district (NSO and ICF, 2017)), we attempted to make the sticker content as clear as possible using images.

1). The prediction is that investment rates should differ across experts and non-experts absent the sticker, but not in the presence of the sticker.²⁶

To classify women as experts vs. non-experts, we could not administer the quality quiz used in the signaling experiment and elicit beliefs about performance on that quiz.²⁷ Instead, we use a proxy informed by the signaling experiment. We had asked six market math questions to the women who participated in the signaling experiment, and elicited their second order beliefs (beliefs about her husband’s beliefs about her performance in the math questions), using the visual handout shown in online appendix figure C.1. These second order beliefs about the market math score are more predictive of the wife’s own perceptions about her market expertise than anything else (i.e., her beliefs about her husband’s beliefs about her market math score are more correlated with her beliefs about her quality quiz score than her actual market math score or her perceived market math score, see online appendix table C.8). Given this finding, we only administered the 6 market math questions to the market experiment sample and elicited second order beliefs. We classify women as follows:

- “*Non-Experts*”: Women with a prior about their husband’s belief about their math score of below 5 (44% of the sample)
- “*Experts*”: Women with a prior about their husband’s belief about their math score of 5 or 6 (56% of the sample)

We estimate the impact of the stickers on the willingness to pay using the following equation:

$$WTP_i = \alpha + \beta_1 NE_i + \beta_2 D_i + \beta_3 E_i + \beta_4 (D \& E_i) + \beta_5 (NE_i \times D_i) + \beta_6 (NE_i \times E_i) + \beta_7 (NE_i \times D \& E_i) + \beta_8' X_i + \mu_e + \delta_m + \epsilon_i \quad (4)$$

where WTP_i is woman i ’s willingness to pay for the good. NE_i is as an indicator that is 1 if the woman believes that her husband would believe she would have a math score below 5, D_i is an indicator that is 1 if woman i is assigned to the donated treatment arm, E_i is an indicator that is 1 if woman i is assigned to the effectiveness treatment arm, and $D \& E$ is an indicator that is 1 if woman i is assigned to both stickers (donated and effectiveness). ϵ_i are Huber-White robust standard errors. We include enumerator fixed-effects μ_e and market fixed effects δ_m , and show estimations with and without adjusting for individual controls X_i .

²⁶To minimize a potential concern that the effectiveness sticker could be interpreted as a marketing ploy by husbands, we purposefully designed the stickers as information leaflets attached to the products externally. This contrasts with traditional advertisements that are commonly integrated in the product packaging.

²⁷There are two reasons for this: (1) the quiz could have created tensions vis-a-vis market vendors, some of which were selling some of the low quality goods in the quiz (since we procured them from local markets); and (2) going through the quiz takes quite some time because lengthy instructions need to be given.

Wife characteristics are broadly balanced by treatment arms, though not perfectly (see online appendix table C.9). We show results both controlling for covariates and without controls, and the results are unchanged.²⁸

6.2 Results

Table 4 presents the results from the market experiment. All results are robust to omitting our vector of controls and controlling for market fixed effects.

The willingness to invest differs substantially by expertise when women cannot easily prove to their husbands that the good is donated or effective. In the control arm, expert wives have an average willingness to pay of MWK 351 and non-expert wives have an average willingness to pay of MWK 279 (-21%, $p < 0.10$, column 3). Neither the donated nor the effectiveness sticker treatment affect the willingness to pay of expert wives, but the donated treatment increases the average willingness to pay of non-expert wives by MWK 120 (+34%, $p < 0.05$) and the effectiveness treatment increases the average willingness to pay by MWK 124 (+35%, $p < 0.05$).²⁹ The effect of the two stickers combined is in the same direction, though somewhat muted, and not significant. Anecdotally, it seems to be because the two stickers combined occupied too much space on the goods and thus made them less attractive.

According to the model, we should see stronger effects among women offered a book (high hiding cost) than women offered the bags (low hiding cost). We present results by the type of good offered in table 5. For simplicity, we pool all stickers into one “sticker arm” since their predicted effects are of the same sign and their observed effects cannot be distinguished from each other in table 4. In line with our model prediction, the investment gap between expert and non-expert wives absent a sticker is much greater when the hiding cost is high than when it is low.

²⁸The most concerning imbalance is that expert wives assigned to the effectiveness sticker treatment have been married longer, hence have larger families, and received a significantly lower transfer from their husband in the previous two months. This is not the case for non-expert wives assigned to that arm. This means that a differential impact of the effectiveness sticker by expertise could possibly be due to these differences, especially due to differences in income. These could not explain differences in impacts of the ‘donated sticker’, however.

²⁹This suggests that at least a fraction of women did not perceive the labels as marketing ploys.

Table 4: Investment experiment: Willingness to pay, by wife's expertise

	(1)	(2)	(3)
Non-Expert	-89.782 (36.868)	-83.491 (37.585)	-71.950 (36.691)
'Donated' Sticker	-32.684 (37.855)	-33.658 (37.627)	-34.968 (37.480)
'Donated' Sticker*Non-Expert	113.829 (55.173)	111.885 (55.872)	120.211 (55.378)
'Effectiveness' Sticker	-31.535 (40.388)	-28.041 (39.560)	-33.970 (38.703)
'Effectiveness' Sticker*Non-Expert	121.058 (59.744)	111.867 (59.137)	124.175 (58.317)
'Donated' & 'Effectiveness' Stickers	-45.036 (39.776)	-46.982 (39.270)	-42.081 (39.157)
('Donated' & 'Effectiveness')*Non-Expert	48.426 (53.765)	47.738 (54.564)	55.670 (53.872)
Mean(Control & Expert)	350.802	350.802	350.802
Observations	675	675	675
Market FE		✓	✓
Controls			✓

Notes: The table shows results from OLS regressions with Huber-White robust SEs. Dependent variable is willingness to pay in Malawian Kwacha. Non-Expert is an indicator that takes the value 1 if the wife reports that her husband has a prior about her math score below 5 (44%). All regressions include enumerator fixed effects. Market fixed effects are dummies for the different markets in which the enumerators recruited married women. Controls include wife's age, education, average income in the last and previous month, risk preferences, math score, as well as the husband's average income in the last and previous month, years married, number of children and household members, and the average transfer from the husband to his wife in the preceding two months.

To more directly test whether non-expert wives have a lower willingness to invest because they are concerned about the dynamics within their household (i.e., their reputation in the eyes of their husband, as opposed to, for example, the eyes of their friends or neighbors), we varied the salience of the husband-wife relationship. Specifically, in half of the sample the BDM was played before the survey and in half of the sample the BDM was played after the survey, i.e., after eliciting the wife's beliefs about her husband's beliefs about her score and after surveying the wife about the previous

two month's transfers and her financial decision-making inside the household. As presented in table 8, the investment gap between expert and non-expert wives absent a sticker is much greater in the salience treatment arm than in the control arm.

Table 5: Investment experiment: Willingness to pay (MWK), by wife's expertise and hiding cost

	Low cost: storage bags			High cost: children's books		
	(N=500)			(N=175)		
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Expert	-84.398 (42.638)	-70.739 (43.264)	-60.593 (41.831)	-189.723 (85.802)	-201.165 (86.877)	-156.137 (85.169)
Any Sticker	-27.706 (37.228)	-23.618 (36.879)	-34.917 (36.937)	-75.234 (60.397)	-79.829 (58.555)	-48.578 (61.766)
Any Sticker*Non-Expert	71.465 (50.930)	60.407 (51.590)	73.982 (50.327)	249.742 (98.013)	262.036 (99.679)	217.482 (97.101)
Mean(Control & Expert)	358.537	358.537	358.537	394.231	394.231	394.231
Market FE		✓	✓		✓	✓
Controls			✓			✓

Notes: The table shows results from OLS regressions with Huber-White robust SEs. Non-Expert is an indicator that takes the value 1 if the wife reports that her husband has a prior about her math score below 5 (44% of the sample). Any sticker is an indicator that is 1 if the wife was assigned to any of the sticker treatments. For fixed effects and controls, see Table 4 notes.

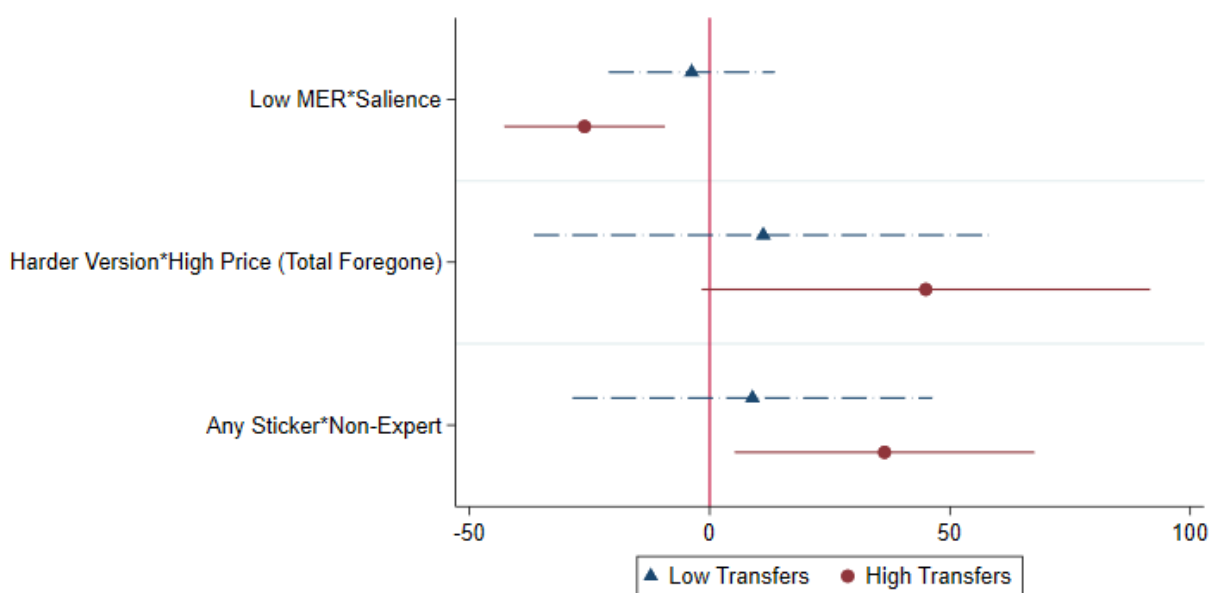
These findings are in line with the idea that women internalize potential reputation costs when making real-life investment decisions. These results have important policy implications as they suggest that women might have a limited ability to experiment with new technologies unless it is ensured that they are able to credibly convey certain information to their husbands.

7 Heterogeneity by Discretionary Transfer Size

The experimental results presented in the preceding sections suggest that husbands' financial transfers respond to their beliefs about their wives' expertise as investors and that wives internalize their husbands' beliefs in their investment and hiding decisions. We now test prediction 4 of the model by assessing heterogeneity in the size of the transfer from husbands to wives in all three experiments. Our model predicts that spouses' behavior should stop responding to the wife's reputation once the reputation has fallen below the cutoff reputation P^* above which discretionary transfers occur. We thus compare experimental results in households in which the wife receives

only “subsistence level” transfers (i.e., transfers for basic household necessities) and households in which the wife receives additional discretionary transfers for investments. In the data, this corresponds to households below or above the median transfer size (as reported by the wives). If we observe significant estimates only in households in which the wife’s reputation is still above the cutoff, and women thus still receive discretionary transfers, we can deduce that our experimental findings reflect real-life concerns about transfers in the household rather than, for example, experimenter demand effect.

Figure 1: Heterogeneity by discretionary transfer size



Notes: The graph shows the coefficients and confidence intervals from OLS regressions with Huber-White robust SEs. Rows 1 and 2 control for enumerator and compensation fixed effects (and version fixed effects for the transfer experiment) as well as wife and husband’s age, education, average income in the last two months, variability of income (whether income is the same in most months or varies a lot), risk preferences, math and raven scores, and years married, number of children and number of household members. Controls are as reported by the husband in the transfer experiment and as reported by the wife in the signaling experiment. Row 3 controls for enumerator and market fixed effects as well as the wife’s age, education, average income in the last two months, risk preferences, math score, as well as the husband’s average income in the last two months, years married, and the number of children and household members. Coefficients are presented as percentage point deviations from the control means.

Figure 1 shows heterogeneity by the mean transfers from the husband to the wife in the previous two months for all three experiments (all coefficients are shown in percentage point deviations from the control mean). The first row plots the coefficients of the effect of the interaction of Low MER*Saliency on dictator game transfers in the transfer experiment. All husbands should use the dictator game to transfer for basic necessities (given the multiplier), and reduce their own private

transfers to the wife afterwards. Only husbands who also transfer for non-necessities investments should react to the salience treatment and reduce their transfers for non-necessities if their wife has low MER. This is exactly what we see: the reduction in transfers in the dictator game to low MER wives in the salience treatment is entirely driven by men who regularly make large transfers to their wives. The second row plots the coefficients of the effect of the interaction of Harder Version*High Price on women's foregone income (foregone participation fee + hiding fee) in the signaling experiment. Again, consistent with prediction 4 of the model, we find that women who receive low transfers from their husbands do not forgo earnings in order to avoid sending a signal. The third row plots the coefficients of the effect of the interaction of Any Sticker*Non-Expert on women's willingness to pay for the goods in the market experiment. Here again, we find that only women who receive high transfers from their husbands exhibit the under-investment behavior predicted by the model. Taken together our results thus provide strong evidence for the external relevance of the experiments as husbands' and wives' behaviors seem to be indeed driven by real-life reputation concerns inside the household.

8 Conclusion

This paper offers a new perspective on some potential dynamics at play between spouses in contexts where women are specialized in household production and at least partly dependent on their husband's income. We develop and test a signaling model in which a woman's access to the household budget varies with her husband's perceptions about her skills as an investor. We show that in order to maintain control over a greater share of the household budget, women experiment too little on average in domains where they have difficulty assessing the productivity of new goods and technologies. What's more, when they do experiment with new goods, they incur costs to hide bad purchasing decisions. Hence our model is able to explain behavior akin to the sunk cost fallacy—using a product even after one has realized it does not have positive returns—within the realm of neoclassical economics.

Three experiments were designed to test specific pieces of the theory. The transfer experiment shows that husbands whose wives made bad market choices in the past transfer less to their wives in a dictator game with multiplier if asked to recall these choices just before playing the game. This is consistent with husbands taking into account the likely return that their wives will be able to obtain when they invest. The signaling experiment shows that women are willing to forgo earnings in order to avoid sending a bad signal about their investment skills to their husbands. Finally, the market experiment shows that these concerns have a direct bearing on a woman's willingness to invest in new technologies. In all three experiments, results are driven by couples in which the reputation of the wife has not yet fallen below the cutoff level - providing additional evidence for the external relevance of our experimental findings.

From a policy perspective, our novel insights about dynamic reputation concerns within the household might help explain the relatively low willingness to pay for high-return investments observed in many programs and experiments targeting women (Cohen and Dupas, 2010; Meredith et al., 2013). Campaigns promoting new technologies or goods could potentially be more successful, and pose a smaller reputational risk to women, if they involved both spouses or ensured that women have the means to credibly convey information about the benefits of the goods to their spouses.

What's more, from an empirical standpoint, the existence of what we coin the "intra-household sunk cost fallacy" suggests that collecting data on *usage* of a given good may not be sufficient to ascertain its value to the household.

Finally, while this paper focused on the husband as principal and wife as agent given the prevailing context of gender inequality, there is no reason why the mechanism would not be completely symmetric in a context where spousal roles are reversed. When both spouses earn equal income, reputation may still matter for the share of the budget that one has control over. For example, a husband who purchased a yogurt maker that was used only twice in the past year may face resistance when he next suggests buying a bread maker. The welfare implications of such dynamics are likely much less stark in contexts where households can afford yogurt makers and bread makers, compared to contexts with limited and unequal consumption such as the one we consider in rural Malawi.

References

- Anderson, Siwan and Jean-Marie Baland**, "The Economics of Roscas and Intra-household Resource Allocation," *The Quarterly Journal of Economics*, 2002, 117 (3), 963–995.
- Andreoni, James and John Miller**, "Giving according to GARP: An experimental test of the consistency of preferences for altruism," *Econometrica*, 2002, 70 (2), 737–753.
- **and Lise Vesterlund**, "Which is the fair sex? Gender differences in altruism," *The Quarterly Journal of Economics*, 2001, 116 (1), 293–312.
- Apedo-Amah, Marie Christine, Habiba Djebbari, and Roberta Ziparo**, "Gender, Information and the Efficiency of Household Production Decisions: An Experiment in Rural Togo," IZA Discussion Papers 12948, Institute of Labor Economics (IZA) 2020.
- Arkes, Hal and Catherine Blumer**, "The psychology of sunk cost," *Organizational Behavior and Human Decision Processes*, 1985, 35 (1), 124–140.
- Ashraf, Nava**, "Spousal control and intra-household decision making: An experimental study in the Philippines," *The American Economic Review*, 2009, pp. 1245–1277.
- **, Erica Field, and Jean Lee**, "Household bargaining and excess fertility: an experimental study in Zambia," *The American Economic Review*, 2014, 104 (7), 2210–2237.
- **, Erica M. Field, Alessandra Voena, and Roberta Ziparo**, "Maternal Mortality Risk and Spousal Differences in the Demand for Children," NBER Working Papers 28220, National Bureau of Economic Research, Inc December 2020.

- Bangladesh Bureau of Statistics and Bangladesh Bureau of Statistics**, “Report on violence against women (VAW) survey 2015,” Technical Report 2016.
- Becker, Gordon M, Morris H DeGroot, and Jacob Marschak**, “Measuring utility by a single-response sequential method,” *Behavioral science*, 1964, 9 (3), 226–232.
- Boltz, Marie, Karine Marazyan, Paola Villar et al.**, “Income Hiding and Informal Redistribution: A Lab in the Field Experiment in Senegal,” *Working Paper*, 2016.
- Cattell, Raymond B**, “Theory of fluid and crystallized intelligence: A critical experiment.,” *Journal of educational psychology*, 1963, 54 (1), 1.
- Chiappori, P.-A.**, “Rational Household Labor Supply,” *Econometrica*, 1988, 56 (1), 63–89.
- Chiappori, Pierre-Andre**, “Collective Labor Supply and Welfare,” *Journal of Political Economy*, June 1992, 100 (3), 437–467.
- Cohen, Jessica and Pascaline Dupas**, “Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment*,” *The Quarterly Journal of Economics*, 02 2010, 125 (1), 1–45.
- Delavande, Adeline, Xavier Giné, and David McKenzie**, “Measuring subjective expectations in developing countries: A critical review and new evidence,” *Journal of development economics*, 2011, 94 (2), 151–163.
- Fu, Chao and Juan Pantano**, “Parental Reputation,” *Working Paper*, 2015.
- Hao, Lingxin, V Joseph Hotz, and Ginger Z Jin**, “Games parents and adolescents play: Risky behaviour, parental reputation and strategic transfers,” *The Economic Journal*, 2008, 118 (528), 515–555.
- Hotz, V Joseph and Juan Pantano**, “Strategic parenting, birth order, and school performance,” *Journal of population economics*, 2015, 28 (4), 911–936.
- Jakiela, Pamela**, “Equity vs. efficiency vs. self-interest: on the use of dictator games to measure distributional preferences,” *Experimental Economics*, 2013, 16 (2), 208–221.
- Jayachandran, Seema**, “The Roots of Gender Inequality in Developing Countries,” *Annual Review of Economics*, 2015, 7 (1), 63–88.
- Kanodia, C, R Bushman, and John Dickhaut**, “Escalation errors and the sunk cost effect - an explanation based on reputation and information asymmetries,” *Journal of Accounting Research*, 1989, 27 (1), 59–77.
- Kenya National Bureau of Statistics and Ministry of Health and ICF**, “Kenya Demographic and Health Survey 2014,” 2017.
- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz**, “Experimental analysis of neighborhood effects,” *Econometrica*, 2007, 75 (1), 83–119.
- Kremer, Ilan and Andrzej Skrzypacz**, “Dynamic signaling and market breakdown,” *Journal of Economic Theory*, March 2007, 133 (1), 58–82.
- Mailath, George and Larry Samuelson**, *Repeated Games and Reputations: Long-Run Relationships* 01 2007.
- Manser, M. and Martin Brown**, “Marriage and Household Decision-Making: A Bargaining Analysis,” *International Economic Review*, 1980, 21 (1), 31–44.
- Meredith, Jennifer, Jonathan Robinson, Sarah Walker, and Bruce Wydick**, “Keeping the doctor away: Experimental evidence on investment in preventative health products,” *Journal of Development Economics*, 2013, 105, 196 – 210.
- Ministerio de Salud Pública y Asistencia Social and Instituto Nacional de Estadística (INE) and ICF**, “Encuesta Nacional de Salud Materno Infantil 2014-2015. Informe Final. Guatemala,”

2017.

- Mobarak, Ahmed Mushfiq, Puneet Dwivedi, Robert Bailis, Lynn Hildemann, and Grant Miller**, “Low demand for nontraditional cookstove technologies,” *Proceedings of the National Academy of Sciences*, 2012, *109* (27), 10815–10820.
- Noldeke, Georg and Eric van Damme**, “Signalling in a Dynamic Labour Market,” *Review of Economic Studies*, 1990, *57* (1), 1–23.
- NSO and ICF**, “Malawi Demographic and Health Survey 2015-16,” Technical Report, National Statistical Office Malawi and ICF, Zomba, Malawi 2017. Available at <http://dhsprogram.com/pubs/pdf/FR319/FR319.pdf>.
- Prendergast, Canice and Lars Stole**, “Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning,” *Journal of Political Economy*, 1996, *104* (6), 1105–34.
- Schaner, Simone**, “Do Opposites Detract? Intrahousehold Preference Heterogeneity and Inefficient Strategic Savings,” *American Economic Journal: Applied Economics*, April 2015, *7* (2), 135–74.
- Smith, Sharon G, Xinjian Zhang, Kathleen C Basile, Melissa T Merrick, Jing Wang, Marcie jo Kresnow, and Jieru Chen**, “The national intimate partner and sexual violence survey: 2015 data brief–updated release,” 2018.
- Swinkels, Jeroen M.**, “Education Signalling with Preemptive Offers,” *Review of Economic Studies*, 1999, *66* (4), 949–970.
- Tauchen, Helen V, Ann Dryden Witte, and Sharon K Long**, “Domestic violence: A nonrandom affair,” *International Economic Review*, 1991, pp. 491–511.

A Empirical Appendix

Table 6: Correlations between reputation measures and transfers from the husband to the wife in the previous two months (MWK) as well as the share of wives who have access to cash and savings

	Avg. transfers in the last two months (MWK)				Access to cash (%)
	(1)	(2)	(3)	(4)	(5)
MER=2	358.419 (1888.604)		329.260 (1887.673)		
MER=3	1241.208 (1788.489)		1194.065 (1791.929)		
MER=4	1789.962 (1772.794)		1672.093 (1795.412)		
High MER		1261.256 (838.000)		1181.115 (851.545)	9.908 (4.035)
High GAR			312.353 (648.896)	398.844 (638.366)	2.067 (2.909)
Control Mean	8902.18	8153.60	8902.18	8153.60	65.05
Observations	1093	1093	1093	1093	1092

Notes: The table shows results from OLS regressions with Huber-White robust SEs. The data is winsorized at 3 SDs (1.7% of the data). MER=2, MER=3, and MER=4 are binary variables that take the value 1 if the woman has an MER of 2 (13%), 3 (31%), or 4 (52%) respectively. High MAR is an indicator that takes the value 1 if the woman has an MER of 3 or 4. High GAR is a binary variable that takes the value 1 if the woman has a General Ability Reputation above the median, and 0 otherwise. Access to cash is an indicator that takes the value 1 if the husband reports that his wife has “access to cash and savings”. The regressions control for wife and husband’s age, education, average income in the last two months (as reported by the husband), variability of income (whether income is the same in most months or varies a lot), risk preferences, math and raven scores, as well as years married, number of children and household members and enumerator fixed effects.

Table 7: Transfer experiment: Effect of reputation salience on amount (%) transferred from the husband to the wife

	(1)	(2)	(3)
Saliency	1.723 (1.736)	2.020 (2.184)	3.141 (2.285)
Low MER	-0.463 (3.376)	-0.127 (3.382)	-0.624 (3.369)
Low MER*Saliency	-9.084 (4.176)	-8.815 (4.192)	-8.894 (4.168)
Low GAR		-2.141 (2.376)	
Low GAR*Saliency		-0.581 (3.020)	
Low Husband GAR			2.108 (2.379)
Low Husband GAR*Saliency			-2.950 (3.011)
Mean (Control & High MER & High GAR)	68.889	70.877	70.877
Observations	1093	1093	1093
Benchmark specification from table 2	✓		
Testing for experimenter demand effect		✓	
Testing for effect of the husband's mood			✓

Notes: The table shows results from OLS regressions with Huber-White robust SEs. Market Expertise Reputation (MER) defined as before. Column 2 tests for experimenter demand effect by assessing the impact of the salience treatment by the wife's General Ability Reputation (GAR). GAR is the normalized mean of the husband's beliefs about the wife's correct answers in a math test, and a raven game. Low GAR is a binary variable that takes the value 1 if the woman has a General Ability Reputation below the median, and 0 otherwise. Column 3 tests whether the salience treatment works through making husbands angrier by assessing the impact of the salience treatment by the husband's General Ability Reputation (GAR). This is the normalized mean of the husband's beliefs about his correct answers in a math test and a raven game. Low Husband GAR is a binary variable that takes the value 1 if the husband has a general ability reputation below the median, and 0 otherwise. All regressions include enumerator, compensation and version fixed effects. Controls include wife and husband's age, education, average income and transfers in the last two months, variability of income (whether income is the same in most months or varies a lot), risk preferences, math and raven scores, and years married, number of children and number of household members.

Table 8: Investment experiment: Willingness to pay (MWK), by wife’s expertise and husband-wife relationship salience

	Control (N=329)			Salience (N=346)		
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Expert	-72.675 (60.330)	-60.950 (62.163)	-59.406 (56.645)	-114.625 (48.082)	-122.298 (46.803)	-107.410 (47.693)
Any Sticker	-33.193 (45.810)	-40.010 (44.358)	-56.384 (44.548)	-42.553 (44.975)	-48.121 (43.423)	-41.495 (44.216)
Any Sticker*Non-Expert	99.642 (71.989)	82.798 (72.401)	117.668 (68.286)	106.902 (57.357)	124.140 (57.665)	122.337 (58.689)
Mean(Control & Expert)	377.885	377.885	377.885	357.143	357.143	357.143
Market FE		✓	✓		✓	✓
Controls			✓			✓

Notes: The table shows results from OLS regressions with Huber-White robust SEs. Non-Expert is an indicator that takes the value 1 if the wife reports that her husband has a prior about her math score below 5 (44%). Any sticker is an indicator that is 1 if the wife was assigned to any of the sticker treatments. Salience is an indicator that is 1 if the BDM was conducted after the survey. All regressions include enumerator fixed effects. Market fixed effects are dummies for the different markets in which the enumerators recruited married women. Controls include wife’s age, education, average income in the last and previous month, risk preferences, math score, as well as the husband’s average income in the last and previous month, years married, number of children and household members, and the average transfer from the husband to his wife in the preceding two months.