Rational Habit Formation: Experimental Evidence from Handwashing in India

By Reshmaan Hussam, Atonu Rabbani, Giovanni Reggiani, Natalia Rigol*

We test the predictions of the rational addiction model, reconceptualized as rational habit formation, in the context of handwashing in rural India. To track handwashing, we design soap dispensers with timed sensors. We test for rational habit formation by informing some households about a future change in the returns to daily handwashing. Monitoring and incentives raise handwashing contemporaneously, and effects persist well after they end. In addition, people are rational about this habit formation: when they anticipate future monitoring, they increase their current handwashing. Average child weight and height increase for all study arms given soap dispensers.
JEL: D04, D83, D91, I12
Keywords: Sex Selection, Marriage Market, Bargaining Power

Habits shape much of how we move in the world. Despite their consideration in economic theory and cognitive psychology (Marshall, 1890; Andrews, 1903; James, 1914; Pollak, 1970; Stigler and Becker, 1977; O'Donoghue, 2001), however, rarely do economists frame behavior change or technology adoption interventions around the development of routine, or habit. We focus on a highly consequential real world setting to test the prevailing economic theory around habit formation. Our study utilizes high frequency time-series data to examine the temporal dynamics of the habit-formation process. It then uses the results to inform policy design for the cultivation of good habits.

Bacterial and viral contamination, resulting in diarrheal disease and acute respiratory infection, kill over two million children per year and stunt the growth of millions more. Handwashing with soap is "the most effective vaccine against childhood infections" (Iyer and Sara, 2005), reducing person-to-person transmission and sanitizing the last point of contact between germs and the mouth (Barker, Vipond and Bloomfield, 2004; Sanderson and Weissler, 1992; WHO, 2009). Despite enormous funding for hand-hygiene campaigns over the last thirty years, however, we know little about how to improve hygiene behavior sustainably. Most interventions find no impact on behavior and/or health (Chase and Do, 2012; Cameron, Shah and Olivia, 2013; Galiani et al., 2016; Null et al., 2018; Luby et al., 2018). The few that do are intensive omnibus interventions (including information, resources, monitoring, and other hygiene and sanitation recommendations) in which multiple mechanisms may be at work (Luby et al., 2005; Bennett, Naqvi and Schmidt,

* Hussam: Harvard Business School, Morgan Hall, 15 Harvard Way, Boston, MA 02163, rhussam@hbs.edu. Rabbani: Department of Economics, Dhaka University, Department of Economics, Social Science Building, University of Dhaka, Dhaka, 1000, Bangladesh, atonu.rabbani@du.ac.bd. Reggiani: Weiss Asset Management, giovanni.reggiani@gmail.com. Rigol: Harvard Business School, 215 Rock Center, 60 N Harvard st., Boston, MA 02163, nrigol@hbs.edu.We are indebted to research assistant Sami Safiullah for his excellent field work and project management. We thank the Birbhum Population Project (BIRPOP) of the Society for Health and Demographic Surveillance (SHDS), a Health and Demographic Surveillance Site in West Bengal, India: in particular, the team of surveyors, survey managers, field monitors, data entry operators, data managers, and the research team. We are extremely grateful to Nan-Wei Gong, who designed the sensors, as well as AQS, who manufactured them. We received guidance from Esther Duflo, Abhijit Banerjee, Frank Schilbach, Ben Olken, Chris Udry, Dean Karlan, Mushfiq Mubarak, Lasse Brune, Hannah Trachtman, David Levine, Vincent Pons, Rafael Di Tella, Matthew Weinzierl, as well as helpful comments from conference and seminar participants at UC Berkeley, University of Chicago, MIT, Yale, University of Michigan, UC San Diego, Georgetown University, Columbia University, Delaware University, the Harvard Population and Growth Center, and Harvard Business School. We are grateful for financial support from the J-PAL Urban Services Initiative, the Weiss Family Fund, the Schultz Fund, USAID, and the Evenson Fund. Experiment registered in the AEA Registry under AEARCTR-0000974. This study received IRB approval from MIT under the COUHES Protocol #1406006477.

2018; Haggerty et al., 1994; Han and Hlaing, 1989).

One feature of handwashing that may explain the difficulty of sustained change is that the behavior must be repeated, and this repetition becomes costly unless routinized. For example, 57% of households in our sample in rural West Bengal, India articulate, unprompted, that they do not wash their hands with soap because *"obhyash nai,"* or "I do not have the habit." The need for repetition is not unique to handwashing. Preventive health behaviors often require routines: water should be treated daily, clean cookstoves used for each meal, medicine consumed at regular intervals, and handwashing engaged in during the same critical moments each day. Most of these behaviors suffer low rates of takeup in the developing world despite their affordability, and neither information provision nor resources appear to generate sustained improvements in such practices (Dupas and Miguel, 2016; Clasen et al., 2014; Kremer and Peterson Zwane, 2007; Banerjee et al., 2010; Chase and Do, 2012; Cameron, Shah and Olivia, 2013). Given their repetitive nature and ties to contextual cues, the psychology literature suggests that such behaviors are ideal candidates for habit formation (Wood and Neal, 2007).

In this study, we examine whether handwashing is indeed a habit-forming activity and whether individuals internalize and respond to its habitual nature. Motivated by the economic theory of rational addiction (Becker and Murphy, 1988), we set up and design an experiment that tests the main implications of this model, overcoming the identification concerns typical to the literature on rational addiction. Our interventions raise handwashing rates by a substantial margin, within eight months translating to improvements in child height-for-age and weight-for-age z-scores of 0.21 and 0.13, respectively.

In order to track the dynamics of habit formation in handwashing, we first develop a novel technology. In partnership with the MIT Media Lab, we designed a time-stamped sensor technology embedded in a liquid soap dispenser, which we then manufactured at scale in China at the cost of approximately $25 USD per dispenser. This technology addresses the challenges faced by all existing handwashing measures: desirability bias (behavior is self-reported or observed by enumerators), subjectivity (cleanliness is subjectively graded by enumerators), noise (metrics are broad and data infrequent), and nonspecificity (presence of barsoap, a common measure, is often due to bathing and laundry rather than handwashing). Our sensor is neither visible nor accessible to households, yielding objective data; it is precise, measuring use at the second level and allowing us to connect observed use with critical times of use; and it tracks the use of liquid soap, which is uniquely associated in our context with handwashing.

In our conceptual framework, habit formation is generated by intertemporally linked preferences in consumption: the more one consumes in the past, the easier consumption is in the present. Intertemporal complementarities imply that front-loaded (i.e. temporary) interventions, which maximize initial takeup, can generate a larger stock of consumption - and thereby greater persistence in behavior - than interventions that are spread over time. This mechanism in turn yields additional predictions given by theory.

In particular, Becker and Murphy (1988), who popularized the theoretical framing of habits (or equivalently, addictive behaviors), posit the theory of "rational addiction." Rationality implies that agents are aware of the intertemporal complementarities of a behavior and foresee their future consumption path accordingly. If a behavior is indeed habit-forming but agents fail to internalize this feature in their consumption decisions, they will underinvest, justifying short-term subsidies to boost usage. Alternatively, if agents are rational habit formers, an intervention that increases the future value of an activity will generate a larger habit stock - and thereby more persistence - than one that raises only the present value of the activity. In its starkest form, if an individual is a rational habit-former, a large, one-time incentive to engage in the future will motivate him or her to engage today (and increasingly so as the future nears) despite the absence of any contemporaneous incentives. Equivalently, front-loaded incentives will generate greater behavioral change by simply being pre-announced. Understanding the nature of the behavior (habitual or not) and how an agent conceptualizes the behavior (rationally or otherwise) is thus important for the optimal design of interventions.

In order to generate habit formation, our experimental design must first increase handwashing rates contemporaneously - a challenge in and of itself as evidenced by existing literature. To do so, we address two critical features of preventive health behaviors. First, the returns to preventive health behaviors, by virtue of being preventive, are not salient. Agents' perceived returns may therefore be lower than the true returns to the activity. In this setting, incentives that present agents with measurable and immediate returns to good behavior may be an effective way to increase initial takeup. Second, preventive health behaviors often exist outside of social norms: given ubiquitously low takeup, individuals face minimal social costs to shirking. In this setting, an intervention that monitors activity and thereby invokes social pressure to engage may activate behavior change.

Our experiment is therefore designed to raise contemporaneous preventive health activity and then test for the presence of habit formation and rational habit formation. We draw from the psychology literature and capitalize on our technology to make handwashing amenable to habituation, using the classic habit loop: a trigger (the evening meal), a routine (handwashing) and a reward (monetary or social incentives) (Duhigg, 2012; Aunger et al., 2010; Neal et al., 2015).

Specifically, we distribute handsoap dispensers with liquid soap and sensor technology to a random subset of households in our sample. Within this group, the experiment has two arms. In the first, we inform households that we are monitoring their activity with the sensor and provide reports on daily handwashing performance (a form of social incentives). In the second, we additionally offer daily financial incentives for handwashing in the form of tickets that are redeemable for household goods. Social and financial incentives are removed after four months, and we continue to track behavior. Persistence in handwashing after the withdrawal of incentives is consistent with handwashing being a habit-forming activity. To test for rational habit formation, we experimentally vary whether agents *anticipate* these interventions: to a subset of households, we announce two months in advance that they should look forward to receiving a monitoring service or extra daily incentives at a future specified date.[1] A present reaction to anticipated changes in future handwashing confirms that individuals are aware of the intertemporal complementarities between performance today and performance tomorrow.

We find that, relative to those who receive only a dispenser, monitoring raises short run handwashing rates by 8.5 percentage points (24%). These higher handwashing rates persist after the withdrawal of the service, suggesting that handwashing is indeed a habit-forming activity. Additionally, we find compelling evidence of rational habit formation among households who anticipated the monitoring service: these households wash 9.3 percentage points (35%) more than their non-anticipating counterparts, with the difference increasing as the date of the monitoring service approaches. It appears, therefore, that households indeed recognize the habit-forming nature of handwashing, and they act upon this knowledge by accumulating habit stock in preparation for a future rise in the value of handwashing.

Adding financial incentives to monitoring further increases handwashing rates. Those who receive financial incentives wash 62% of the time, relative to 45% for those who receive monitoring and 36% for those who receive only the dispenser. After incentive withdrawal, higher handwashing rates persist, substantiating evidence on the habitual nature of the activity. However, these effects are not mirrored on the intensive margin of financial incentives, along which we test for rational habit formation: those who experience an *increase* (in particular, a tripling) in financial incentives wash only 5.1 percentage points (9%) more than their standard incentive counterparts, suggesting

[1]Note that *all* households are notified that the future monitoring service or incentive boost is a possibility at the beginning of the experiment. They are told that resources are limited, so whether or not they should anticipate (receive) these future boosts will be determined by lottery. If households use the implementing organization's incentive scheme as a signal of the true returns to handwashing, this lottery method equalizes expected value of handwashing *as judged by the implementing organization* across anticipating and non-anticipating households.

rapidly diminishing marginal returns to financial incentives. This rate decays to the standard incentive level soon after incentives are withdrawn. We furthermore find no anticipating reaction to the tripling of incentives: households anticipating this change wash no more than their non-anticipating counterparts.

Our results are consistent with the key predictions of a model of rational habit formation. When faced with a future increase in incentives, households appear not to invest in accumulating hand-washing stock to 'prepare' given the low contemporaneous benefit or prospects of habit formation associated with the change. On the other hand, households invest considerably in accumulating stock for an intervention with significant contemporaneous and persistent bite, as evidenced by the monitoring setting.

Lastly, we examine child health outcomes to establish the causal link between handwashing and child health. We find strong effects on child health, confirming that handwashing alone has substantial returns in resource-poor settings. Children in households that received a dispenser and soap report 38% fewer days of loose stool (a proxy for diarrhea episodes) and 16% fewer days of acute respiratory infection (ARI) eight months after distribution (intent-to-treat estimates).[2] These effects rise to 72% fewer days of loose stool and 30% fewer days of ARI when we examine the impact of the treatment on the treated, where 'treated' is defined as those who self-report regularly washing their hands at the eight-month mark.[3] These reductions in morbidity translate to significant improvements in child weight-for-age and height-for-age: treated children under sixty months experience a 0.13 standard deviation increase in their weight-for-age and a 0.21 standard deviation increase in their height-for-age eight months after dispenser distribution. The magnitude of these effects is comparable to recent nutrient supplementation interventions of significantly greater intensity and duration (Null et al., 2018; Luby et al., 2018).

This study makes five contributions. First, to our knowledge, this is the first field experiment designed to test for rational habit formation. The vast majority of the literature, all of which explores bad habits such as smoking and alcohol consumption, rests in favor of rational addiction (Becker, Grossman and Murphy, 1991; Chaloupka, 1991; Cameron, 2000; Baltagi and Griffin, 2002; Gruber and Köszegi, 2001). The classical empirical test of rational addiction regresses present consumption on past and future consumption and other demand shifters, instrumenting for the lag and lead of consumption using the lag and lead of prices:

$$\mathrm{c}_t = \theta \mathrm{c}_{t-1} + \beta\ \theta \mathrm{c}_{t+1} + \delta \mathrm{p}_t + \epsilon_t$$

where a positive coefficient θ is evidence of addictiveness and a positive coefficient $\beta\theta$ is evidence of rational addiction (Becker, Grossman and Murphy, 1994).

However, Auld and Grootendorst (2004) describe the implausible variation in discount rates, unstable demand, and low price elasticities implied by such literature. The authors demonstrate how high serial correlation in the prices of the commodity of interest and endogeneity in the price instruments (where price changes can be due to a supply shock or a demand shock) can yield a positive coefficient on future consumption for non-addictive goods such as milk that is in turn incorrectly interpreted as evidence of rational addiction.

Gruber and Köszegi (2001) make substantial progress on several of these empirical concerns. Recognizing the implausibility of consumer knowledge of far-future price changes, they employ announced tax increases to proxy for exogenous future shocks to consumption. The use of taxes further eases the concern of price endogeneity, as changes in tax rates are more likely to be due to supply side shocks. The potential endogeneity of taxes to cigarette demand is then alleviated

[2]We cannot reject that treatment effects are statistically equivalent across the sub-treatment arms of incentives, monitoring, and dispenser-only, so we report pooled estimates here.

[3]This measure is correlated with actual dispenser use (correlation of 0.48), which we cannot employ in an IV regression since our comparison group of pure control households do not have dispenser use data.

with the inclusion of state-specific time trends. The work to date offers the most carefully identified empirical estimates of rational addiction in cigarette consumption. However, although the announced tax rate change is a significant improvement upon previous work, there is no way to verify whether consumers are aware of the future tax rate, and the likelihood is low given the year or more between the observed consumption decision and the tax enactment. Ideal cigarette demand data is a further challenge: the authors employ sales data, which is vulnerable to hoarding behavior, and consumption data for women who have just given birth, a selected sample for whom underreporting, particularly in times when sentiment against smoking is pervasive, may be high. Importantly, these challenges are directly tied to the nature of the non-experimental, aggregate time-series data available in existing empirical scholarship.

The field experiment nature of our study addresses each concern above. Our design allows us to: (1) impose price changes exogenously, avoiding endogeneity between prices or tax rates and consumption; (2) explicitly announce future prices so consumer knowledge is assured; (3) avoid concerns of differential time trends given randomization; (4) avoid endogenous misreporting using our objective measurement device; and (5) avoid the implications of serial correlation in commodity prices as we impose prices exogenously and randomization permits us to compare outcomes across groups rather than over time. Finally, in a literature that has thus far explored only negative behaviors, this is the first study to examine rational habit formation in the context of *good* habits, a key feature of preventive health behaviors.

Our second contribution is the advancement of the measurement of habit formation, distinct from the test of rational addiction. Existing literature typically equates habit formation with long run persistence of temporary interventions. However, persistence can be due to multiple mechanisms: the purchase of a technology that changes the production function, the process of learning more about an activity such that one updates her perceived costs or returns to engaging, the amplification of one-time behaviors through social interactions, or the accumulation of consumption stock (Allcott and Rogers, 2014; Fujiwara, Meng and Vogl, 2016). Individual-level habit formation is driven only by the last of these. Importantly, our evidence of rational habit formation can be attributed only to changes in future consumption stock, because we experimentally vary only the future value of handwashing behavior, not that of health returns. Paired with an experimental design that includes a dispenser experimentation period across all arms and evidence that handwashing behavior does not vary with the size of child health returns, we can rule out each alternative mechanism and isolate habit formation as the driving force behind persistence in handwashing.

Third, by identifying the marginal impacts of financial incentives, social incentives, and dispenser and soap provision, this study sets an important precedent for the design of public health campaigns, which regularly pool interventions and are unable to disentangle the causal effects of each, often theoretically distinct, dimension of the program. For example, in Luby et al. (2005), a study oft-cited as the hallmark of a successful handwashing campaign, community volunteers visit households twice weekly, deliver soap, instruct and monitor households' handwashing practices, and advise households on other hygiene and sanitation behaviors. While the intervention was remarkably successful at sustainably reducing child diarrhea and respiratory infection, the study is unable to identify which aspect of the intervention led to the health improvements. Beyond distinguishing the effects of resource provision from incentives, our study further addresses mechanism conflation in economic incentive interventions: in any setting in which a conscious principal rewards an agent for performance, a material incentive is a sum of (1) the material reward and (2) social monitoring and feedback. Particularly in contexts where returns are not salient and there are no social costs to shirking, it is important to estimate the impact of each mechanism alone.

Fourth, our data quality is unprecedented within the hygiene and sanitation literature. The objective, high-frequency data of the dispenser sensors allows us the first opportunity to design an experiment which disentangles the various behavioral barriers behind low handwashing takeup. This time-stamped data is also rare in the broader literature of adoption of preventive technologies. It complements the recent collection of energy conservation studies in developed countries that utilize household-level meter data to examine how various informational interventions affect household energy consumption (Allcott and Rogers, 2014; Ito, Ida and Tanaka, 2015; Jessoe and Rapson, 2014; Allcott and Kessler, 2019). Importantly, these studies have been unable to link a specific behavioral change to reduced consumption, whereas the sensor data and design of the present study permit a direct link between increased dispenser use and handwashing with soap during the evening mealtime.

Finally, this study offers the first *treatment on the treated* estimate of the impact of handwashing on child health. In a literature that is plentiful in health impacts of zero, occasional in impacts that are positive yet unable to identify the cause of improved health, and scarce in causal estimates which still say nothing of the ratio of input (handwashing) to output (health), this study offers a significant step forward in establishing the magnitude of impact that handwashing alone can have on health. This helps us build a more precise production function of child health as it relates to preventive behaviors in low-resource settings, which is essential for the more efficient allocation of research and policy funds.

The remainder of the paper proceeds as follows. Section II outlines the conceptual framework motivating our experimental design; Section III describes the study sample and experiment; Section IV specifies our outcomes of interest and the empirical strategy; Section V presents results on handwashing behavior; Section VI presents results on child health; and Section VII concludes.

I. Conceptual framework

Our framework for habit formation builds upon the seminal work of Becker and Murphy (1988) on rational addiction. They and others in their spirit have focused on characterizing and testing the implications of rational addiction in the context of bad habits. We expand to the context of good habits, of which handwashing with soap before mealtime is our focus. Substantively, this manifests as a shift from an activity in which the user experiences positive gains in the present but incurs costs in the future to an activity in which the user incurs costs in the present but experiences positive gains in the future. This model is formalized in Section 2.1. Throughout our discussion, we use 'addiction' and 'habit formation' interchangeably, as their underlying mechanism of intertemporal complementarities is identical.

Intertemporal complementarities in the utility from consumption are an intrinsic property of a habit. *Rational* habit formation (à la Spinnewyn (1981), and conceptually equivalent to what Becker and Murphy (1988) term rational addiction) is the recognition of these properties: a rational habit former is one who internalizes the habit-forming nature of the activity and chooses to engage conditional on this knowledge. The key tradeoff that a rational habit former faces when choosing whether to engage in a good habit is therefore between the drop in utility from consumption today and the increase in long-run utility from the accumulation of the stock in the addictive good.

A. *Model of rational habit formation*

We present a model of rational habit formation for positive behaviors, adapted from O'Donoghue (2001)'s discrete time exposition of rational addiction for bad habits.

Consider a discrete time model with periods 1,...,T. In each period, an agent can wash her hands before dinnertime such that consumption $w_t = 1$, or refrain from handwashing such that $w_t = 0$. Define k_t as the 'habituation level' of the agent in period t:

$$k_t = \gamma k_{t-1} + w_{t-1}, \gamma \in [0, 1) \tag{1}$$

Habituation is a recursive function which is dependent on the agent's habituation to handwashing in the previous period, k_{t-1}, the level of decay the behavior is subject to, γ, and whether the agent washed in the previous period, w_{t-1}.

Define the agent's instantaneous utility function in period t as

$$u_t(w_t, k_t) = \begin{cases} \alpha - [x_t - \sigma k_t] & \text{if } w_t = 1 \\ 0 & \text{if } w_t = 0 \end{cases} \tag{2}$$

where α is the health benefit from washing, x_t is the net cost associated with handwashing before dinnertime (in time, effort, attention), and σ is the ease of washing attributable to habituation. Handwashing is habit forming: the more one has washed her hands in the past, the greater her desire to wash at present ($\frac{\partial u_t}{\partial k_t} > 0$, or $\sigma > 0$).

We wish to maximize the net period-t instantaneous utility of washing:

$$\begin{aligned} u_t(k) &= u_t(1, k) - u_t(0, k) \\ &= \alpha - x_t + \sigma k_t \end{aligned} \tag{3}$$

A myopic agent will only wash her hands today if the marginal cost of washing is outweighed by the health benefit and the benefit from the ease of washing generated by habituation (σ). The more she has washed in the past, the greater the impact of habituation on the marginal utility of consumption and the more likely she is to wash in the present. This intertemporal complementarity in consumption is the essence of habit formation.

$$x_t < \alpha + \sigma(\gamma k_{t-1} + w_{t-1}) \tag{4}$$

What levers can be shifted to generate a positive desire to wash? For agents who have not yet accumulated handwashing stock (most households in our setting), σ offers no leverage, because $k_{t-1} = 0$ and $w_{t-1} = 0$. To facilitate the accumulation of stock, we must focus first on lowering the net cost of handwashing, x_t. If sufficiently lowered, an agent will wash; she thereby enters period $t + 1$ with a $k_t > 0$. If the cost is lowered for a sufficient number of periods, the agent will accumulate enough consumption stock such that, even absent the subsidy on cost, the desire to engage will be positive. In a setting of habit formation, subsidies need only be temporary to generate long run behavioral change.

We actualize the reduction in net cost x_t in two ways. Our first intervention subsidizes the cost of daily handwashing by providing daily financial incentives for good behavior. These incentives are provided by field workers who are members of the communities they survey. In such a context, where incentives are directly linked to discrete units of behavior and a sentient principal is providing the incentives to the agent (a setting hardly unique to this study), incentives are implicitly a sum of (1) financial rewards, (2) feedback on behavior, and (3) the social pressure of being watched and held (financially and/or socially) accountable. The latter mechanism can be conceptualized as increasing the cost to shirking. In order to disentangle the relative importance of the financial reward from the interpersonal cost, we implement both an incentives intervention and an intervention of feedback sans financial reward, which we term 'monitoring.'[4]

Having generated a positive amount of handwashing stock, these interventions can be complemented with an environment which facilitates maximum retention of stock. For example, we can maximize the impact of habituation, σ, by framing the behavior as part of a habit loop: the hand-

[4]This framing of financial and social incentives as two means of shifting the net cost of handwashing was established prior to, and directly motivates, the experimental design. We note that we cannot disentangle the value of the feedback itself from that of social accountability; to do so would require more sophisticated technology around the sensor, similar to a step-counter for tracking exercise. However, the sentient principle and agent is a natural setting which activates both mechanisms.

washing routine can be supported on the front end by the trigger of mealtime and the back end by incentives or monitoring feedback. Given our limited sample size, we choose not to experimentally vary the type of trigger administered (it remains the dinner mealtime for all households), but do vary the feedback by providing households with no feedback, monitoring feedback only, or additionally daily incentives for handwashing (described in more detail in Section 3.3).[5]

Thus far we have considered the instantaneous utility function in a given period which an agent faces for habit-forming behaviors. In a world where agents are forward thinking, the long run utility function is as follows:

$$U_t(k_t) = \max_{w_t} \begin{cases} [\alpha - x_t + \sigma k_t] + \delta U_{t+1}(\gamma k_{t+1}) & \text{if } w_t = 1 \\ \\ \delta U_{t+1}(\gamma k_t) & \text{if } w_t = 0 \end{cases} \tag{5}$$

where $\delta \leq 1$ is the agent's discount factor. A rational habit former is one who *recognizes* the intertemporal complementarities in her utility from consumption. If she is aware that her stock of handwashing in the past affects her likelihood of engaging today, then she is similarly aware that her likelihood of engaging in the future will be affected by her engagement today. Therefore, if an exogenous shock, such as a drop in the future cost of handwashing, changes her likelihood of engaging in the future, she should update accordingly her likelihood of engaging today.

In summary, the model yields the following testable implications.

1) **Incentives:** $\frac{\Delta w_t}{\Delta x_t} \leq 0$. Reducing the cost of handwashing (by increasing the value of handwashing) raises handwashing rates.

2) **Monitoring:** $\frac{\Delta w_t}{\Delta x_t} \leq 0$. Reducing the cost of handwashing (by increasing the cost of shirking) raises handwashing rates.

3) **Habit formation:** $\frac{\Delta w_t}{\Delta k_t} \geq 0$. A rise in past handwashing rates increases current handwashing rates.

4) **Rational habit formation:** $\frac{\Delta w_t}{\Delta w_{t+j}} \geq 0$ with $j \geq 2$. A rise in future handwashing rates w_{t+j}, for example through an exogenous drop in the future cost of handwashing x_{t+j}, increases current handwashing rates.[6]

5) **Health:** $\alpha \geq 0$. Handwashing generates positive health benefits.

[5]Our framework abstracts from three features the reader may find in popular models of rational addiction: first, we eliminate health *internalities*, or the multiplicative role of past consumption on present health benefits. Because we do not vary the magnitude of the health returns across treatment arms (which one may have done if public health information were randomized, but we chose to keep this constant across all arms), we do not find this dimension of the framing instructive. Second, we do not incorporate present-biased preferences into our model. This is simple to do following Gruber and Köszegi (2001) or O'Donoghue (2001)'s discrete time exposition of the same, but unnecessary for our setting: the goal of the experiment is to capture whether agents are forward thinking at all, not whether they are fully forward thinking or "100% rational" (O'Donoghue, 2001). As long as an agent is not 100% myopic, we will detect behavior consistent with [some amount of] rational habit formation. Finally, because we do not vary the trigger, or cue, to handwashing across treatment arms, we do not complicate our framework with cue-centered models of habit formation such as that of Laibson (2001); we stress however that the motivation behind such models is precisely the reason we focus on the singular activity of handwashing before dinnertime, which we encourage across all households. To be sure, we cannot distinguish the role of financial incentives or monitoring from that of cues, since both types of incentives are attached to dinnertime washing.

[6]Note that each of these implications are subject to specific thresholds. Consider the myopic model of habit formation. A reduction in x_t to x'_t will only shift an individual from not washing to washing today if $\Delta x_t = x_t - x'_t > x_t - [\alpha + \sigma k_t]$. Likewise, for the forward thinking agent, rational habit formation will only shift an agent from not washing to washing today if $\Delta\delta U_{t+1}(\gamma k_{t+1}) = \delta U_{t+1}(\gamma k_{t+1}) - \delta U_{t+1}(\gamma k'_{t+1}) > \delta U_{t+1}(\gamma k_t) - \delta U_{t+1}(\gamma k_{t+1}) - [\alpha - x_t + \sigma k_t]$.

II. Experimental Design

A. Study sample and context

Our sample population is made up of 2,943 peri-urban and rural households containing 3,763 children below the age of seven across 105 villages in the Birbhum District of West Bengal, India. Table 2 presents sample means for a host of household, mother, and child characteristics, as well as measures of the mother's hygiene knowledge and practice at baseline, separately for treated and control households. The average mother is just above 31 years old with six years of education. Respondents know a substantial amount regarding hand hygiene: 95% are aware that soap cleans hands, and 79% articulate without prompting that soap cleans germs. However, hygiene practice is poor. Despite more than 96% of respondents reporting that they rinse their hands with water before cooking and eating, only 8% report using soap before cooking and 14% before eating. This failure to use soap cannot be due to lack of soap availability: 99.8% of households report having soap in the home.

Our partner organization, the Society for Health and Demographic Surveillance (SHDS), is a public health organization with a strong presence in the Birbhum District. SHDS surveyors had been visiting all households in our sample twice monthly for one year prior to this study's baseline in order to collect child health data, a practice that we continued for the duration of the present study.

B. Dispenser and soap features

We employed a standard wall-mounted dispenser as depicted in the top picture of Appendix Figure A1, which was outfitted with a time-stamped sensor. Soap was loaded in a one-liter plastic container inside the dispenser and refilled as needed throughout the course of the experiment during the surveyors' biweekly visits. The sensor module is fitted between the container and the soap spout, as shown in the bottom picture of Appendix Figure A1. The circuitry is protected by a waterproof casing, an essential feature for the oft-wet environment of West Bengal. Each push of the outer black button is registered in the sensor to the nearest second. The unit is powered by a rechargeable battery; this was essential given the lack of electricity in many of our rural households. Each soap dispenser cost approximately $4 USD, and each sensor module cost approximately $25 USD at a quantity of 1200 pieces; this cost drops sharply with higher production given the substantial fixed cost of designing the mold for the waterproof casing. This is the first time-stamped sensor technology to be designed for the purpose of handwashing in outdoor, off-grid environments and successfully implemented at scale.

The dispenser was installed near the dining space or water station as chosen by the household. Appendix Figure A2 depicts a typical setting for the dispenser: families usually eat on the veranda. We chose a wall-mounted dispenser because (1) tabletop dispensers were at risk of theft, and (2) a stable 'handwashing station' made it easier for households to remember to wash, potentially enhancing the physical trigger in the habit loop. The dispenser was positioned at a height reachable by young children as shown in Appendix Figure A3.

Identifying an appropriate soap likewise required extensive piloting, which revealed that households preferred: (1) unscented or lightly scented soap that would not interfere with their eating experience, and (2) soap that lathered easily. We thus chose a foaming soap with a light scent approved by pilot households. We preserved some scent as the olfactory system is a powerful sensory source of both memory and pleasure and thus easily embedded into the habit loop (Duhigg, 2012). The foaming soap was both a more pleasant experience for the user and required less water to lather and wash off; although water availability was not a constraint in our setting (where water from shallow tubewells is readily available and regularly used before eating), it is a useful feature for spaces that face water scarcity.

C. *Timeline and treatment groups*

Figure 1 provides a map of all treatment arms and the time-contingent randomization process. Henceforth, treatments associated with social incentives will be referred to as "monitoring", and those associated with financial incentives will be referred to as "incentives."

The duration of the experiment, from baseline to the final round of handwashing data collection, was August 2015 to March 2017. The randomization was conducted in three stages. First, the 105 sample villages were randomized into Monitoring Villages (MV) and Incentive Villages (IV). Interventions across the villages were rolled out in parallel. Households in MV were randomized into two groups: (MV0) control and (MV1) dispenser only. Households in IV were likewise randomized into two groups: (IV0) control and (IV1) dispenser + incentive. Recall that receiving financial incentives implicitly involves receiving feedback and monitoring.

Monitoring and incentive assignments were first randomized at the village level in order to limit the scope for inter-household tension: surveyors expressed concern that control households would be angered if they had some neighboring households who received a dispenser and others who received a dispenser and incentives. It would be easier to justify the interventions through the limited resources lottery framework if all dispenser-receiving households within a village received a consistent package of goods (i.e. the dispenser either always came paired with incentives or never did). This initial-level randomization at the village level also allowed us to cleanly conduct two parallel rational habit formation experiments (one using monitoring, the other using incentives), eliminating concern that households anticipating a future change in services may be confused regarding the type of future change (in other words, households anticipating a future increase in monitoring could not mistakenly also expect future incentives, since they did not live amongst any incentivized households).

At rollout, all households received a basic information campaign regarding the importance of washing hands with soap, especially prior to eating. They also received a calendar with the SHDS logo as a token of appreciation for participation. They were notified that they would be visited biweekly for one year to collect information on child health and (for those who received dispensers) check and replenish soap supplies.

The remainder of the randomizations were conducted at the household level, with households in monitoring villages and those in incentive villages experiencing a parallel evolution in treatments over time. All arms were pre-specified prior to the baseline survey[7]. Each treatment arm is described in detail below.

(MV0 and IV0) Control - 1785 HHs: Households were given a simple lecture on the importance of washing hands with soap, with stress placed on the responsibility of the mother to do so and encourage her household to do so for the sake of her children's health. They also received a calendar as a token of appreciation.

(MV1) Dispenser - 130 HHs: Households were given a soap dispenser. They were informed that there was a switch inside the dispenser that, if turned on, would track their behavior. SHDS wished to offer a monitoring service to the households in which handwashing would be reported biweekly and tracked on their calendar. Because resources were limited, the service would be administered by lottery. If they did not get selected, their switch would not be turned on, their

[7]A pre-analysis plan is published on the AEA RCT Registry at https://www.socialscienceregistry.org/trials/974/history/7386. Two deviations from the plan should be noted. First, we do not control for willingness-to-pay for the monitoring service (a measure we elicit at baseline from all those in MV1) in our regressions because field experience made clear that the elicited valuations are largely noise. Results are robust to controlling for this value and can be provided upon request. Second, although we intended to cross randomize all dispenser arms with a reminders intervention (involving small alarm clocks placed on top of the dispensers), we never received the clocks due to import complications between Shenzhen and Kolkata, and thus had to eliminate this part of the experiment.

behavior would not be monitored, and no feedback would be provided.[8] Note that in practice, all switches were turned on from the beginning, so this arm involves deception.

(MV3) Anticipated monitoring - 233 HHs: Two weeks after dispenser distribution, these households were informed that they had been selected in the lottery: the internal switch would soon be turned on, and the device would record the time and frequency with which the household washed their hands with soap.[9] The surveyor would be observing this data every two weeks and would provide the household with a biweekly report of their behavior, particularly around their evening mealtime, marking the household's calendar in the presence of the mother. This arm can therefore be regarded as a combination of information and feedback, third-party monitoring, and self (or parent-child/intrahousehold) monitoring. The service would begin two months after dispenser distribution on a date circled clearly by the surveyor on the household calendar and written on a sticker attached to the dispenser. This upcoming date was re-announced at each proceeding surveyor visit to ensure comprehension.

(MV2) Unanticipated monitoring - 119 HHs: Two months after dispenser distribution, these households were surprised with an identical monitoring service to those in MV3, effective immediately.

(IV1) Incentives -191 HHs: At the point of dispenser distribution, these households were informed that there was a switch in their dispenser which, when on, tracked the frequency and time of use; and that this switch was on and their behavior was being tracked. They were then given a small coin purse and told that they would receive one ticket for every day in which the device was active during their stated dinnertime, which they should accumulate in their purse. These tickets could be exchanged for various household and child prizes as detailed in a prize catalog.[10] These incentive payments would last for four months. Households were also told that SHDS anticipated receiving additional funding from the government for the project in the near future, at which point SHDS hoped to increase the reward for handwashing by three-fold. Because the future funds were limited, households would be entered into a random lottery to see who would receive the future increase in reward. They would be notified of the results of this lottery within two weeks.

(IV3) Anticipated triple incentives - 310 HHs: Two weeks after dispenser distribution, these households were informed that they had been selected in the lottery for the incentive boost and could soon expect to receive triple the number of tickets for every day in which the device was active during their stated dinnertime for thirty days. The boost would begin two months after dispenser distribution on a date circled clearly by the surveyor on the calendar and written on a sticker attached to the dispenser. As in the monitoring scenario, this date was re-announced at each proceeding surveyor visit to ensure comprehension.

(IV2) Unanticipated triple incentives - 179 HHs: Two months after dispenser distribution, these households were surprised with an identical incentive boost to those in IV3, effective immediately.

It was important that we provided all [incentivized] households with an incentive from the start of the experiment in order to establish (1) an understanding of the nature of the incentives such that households could easily estimate the value of a future boost in incentives and (2) trust between

[8]These lotteries were publicly announced in order to equalize the expected value of the monitoring across receiving and non-receiving households; it preempted the possibility that a household would update its valuation of handwashing because, for example, the provision of an additional service was a signal that they should value the behavior more.

[9]Households could choose whether or not they wanted to receive this program; in practice, all selected households chose to accept it. This was also true for MV2 households.

[10]The ideal incentive requires three conditions: (1) the incentive must be divisible; () the daily amount offered must be sufficiently high to induce behavioral change on a daily basis, which is key to habit formation; and (3) the marginal value of the units accumulated as the process of habit formation continues must also remain sufficiently high to continue inducing behavioral change. Tickets exchanged for goods satisfies all three conditions while also offering flexibility in the types of goods that a household may find appealing. Prizes were selected to focus on child health and schooling and adult household goods.

the surveyors and the households that a claimed future increase would indeed by fulfilled.[11] These necessities preclude a test of rational habit formation on the extensive margin for incentives (which would make them more directly comparable to the monitoring experiment); however, this design most closely mimics the existing literature on rational addiction, all of which examines future intensive-margin price changes on current behavior.

Sample sizes were determined around a few constraints: (1) the partner organization had a pool of 2,947 households in their available sample; (2) budget constraints permitted the production of only 1,400 handsoap sensors and dispensers, of which approximately fifteen percent were reserved as backup (given failure rates in pilots); (3) we aimed to sample more heavily in incentive villages as we anticipated smaller effect sizes on rational habit formation from an intensive-margin change in incentives than from the extensive-margin change in monitoring. Finally, data collected in the experiment can be found in Hussam et al. (2020).

D. *Identification of effects*

The effect of receiving the dispenser and soap alone is captured in the comparison of households in MV1 to MV0.

A higher takeup of handwashing behavior in MV3 relative to MV1 and IV3 relative to IV1 (*before* the price change) demonstrates the presence of rationally habit forming behavior: households who increase take-up today due to an increase in the future value (or decrease in cost) of handwashing must recognize that higher take-up today will accumulate greater habit stock over time, making it easier to reap the benefits of the future rewards to the behavior. q A zero difference in take-up between households in MV3 versus MV1 and IV3 versus IV1 prior to the price change could be due to three reasons: (1) households either fail to recognize the habit-forming nature of the activity and/or are not sufficiently forward-looking, precluding rational habit formation; (2) the future change in the value of handwashing was not sufficiently compelling to induce behavioral change, even for forward-looking individuals; or (3) handwashing is not a habit-forming activity. The second possibility is eliminated if households do indeed respond to the price change (i.e. the tripling of tickets or monitoring service provision) when it is enacted. This contemporaneous effect can be identified by comparing households in MV1 to those in MV2 and households in IV1 to those in IV2 after the price change, as the only difference between these sets of households is the price change itself, with no behavioral response to anticipation. This comparison gives us the pure contemporaneous effect of the incentive boost or the monitoring service on handwashing behavior.

The third possibility is eliminated by comparing persistence in behavior across all arms after the withdrawal of all interventions. For households in arms IV1, IV2, and IV3, all incentives [and implicitly, monitoring] services were discontinued approximately two months after the price change. For households in arms MV2 and MV3, all monitoring services were discontinued approximately four months after their introduction.[12] In practice, households were informed that the switch in their machine had been "turned off," that surveyors would no longer be observing their behavior but would continue to visit monthly to collect child health data, and that surveyors would no longer provide reports on household handwashing performance (nor tickets for incentive households).[13] A

[11] These were lessons learned from our pilot, in which we provided future incentives on the extensive margin, and it was clear that households did not understand what incentives meant nor trust that we would provide them until the future date of change arrived and we delivered the tickets.

[12] This difference in date of discontinuation was implemented to equalize the exposure of households to each treatment, since incentive households had already been receiving incentives for nearly two months prior to the price change.

[13] As is true for MV1 (dispenser-only) households as well, this practice of informing households that the switch in the machine was "turned off" constitutes deception. The practice was cleared by IRB boards at both MIT and IFMR prior to implementation and was permitted given the scientific value and significant policy relevance of the lessons learned. In particular, this practice allows us to estimate the effects of (1) third party monitoring and feedback, yielding a measure of the extent of bias in typical observational outcome measures used in these studies as well as a measure of the role that monitoring effects may play in the cultivation of social norms; and (2) persistence after the withdrawal of interventions, yielding a measure of the sustainability of

comparison of each treatment arm to MV1 households, who were never exposed to any interventions beyond the provision of the dispenser and soap, quantifies the extent to which a handwashing habit was formed due to the temporary incentives or monitoring interventions.[14] We track household handwashing behavior for fourteen months after rollout.

By maintaining the same incentive stream across both groups, a comparison of MV3 to MV2 and IV3 to IV2 over the course of the experiment after the price change allows us to identify the effect of forward looking, rationally addictive behavior on habit formation (conditional on finding evidence of rational habit formation prior to the price change). In other words, a long term comparison of take-up between the 3 and 2 groups demonstrates whether forward-looking behavior in fact facilitated the formation of the handwashing habit.

III. Methods

A. *Outcomes of Interest*

Our primary outcomes of interest encompass behavioral change in households and child health. We capture behavioral change through recorded dinnertime-specific daily handwashing rates and recorded total daily handwashing rates. Note that sensor measures of handwashing rates could only be collected for those households with dispensers, so we do not have data from the pure control households on these metrics. We therefore supplement these with alternative measures of hand hygiene commonly employed in the literature. We collect child health data in the form of self-reported biweekly incidence of child diarrhea and respiratory illness and anthropometric measures of height, weight, and mid-arm circumference. Each is defined in detail below.

A. Household handwashing behavior

Handsoap dispenser data was collected every two weeks during surveyor visits. Although it was not possible to identify the identity of the user at any given press, we proxy for separate users by collapsing presses that happen two or fewer seconds apart into a single press. In other words, if the device is used in seconds 34, 35, 37, 45, and 46, the first three presses are considered a single use by one household member and the latter two presses as a single use by another member. Though not exact, observations from pilots elucidated that users press several times in quick succession and rarely return for more soap during a single handwashing event, since the water source (usually a bucket right outside the front porch) is not within reach of the dispenser (unlike the familiar setting of sink, soap, and running water common to more developed contexts).

Mealtime-specific handwashing rates are calculated as the total number of 'individual' uses in the interval of 90 minutes before and after the household's reported start of the evening meal time. If a family reported eating dinner every day at 8:00 PM, for example, this outcome would be the sum of all individual presses observed between 7:00 PM and 8:30 PM.

Binary use at mealtime is derived from the above and is a binary variable which equals one if at least one 'individual' use was observed in the dinnertime interval. This is the outcome by which we determine calendar markings and tickets earned, and therefore our primary outcome measure of handwashing at dinnertime. All results presented in the paper utilize this outcome unless specified otherwise. Results are robust to using the continuous measure and can be provided upon request. [15]

the interventions and the habit-forming nature of handwashing.

[14]We equate persistence to habit formation under the assumption that persistence is driven purely by the increase in consumption stock accumulated through the interventions, not through the acquisition of a technology that shifts households onto a new hand hygiene path or through learning about the activity or about its returns. These are not trivial assumptions, and we address each in detail in Sections 5.2 and the Appendix.

[15]We choose this binary measure as our preferred measure of "proper" handwashing because we wanted to minimize Type II

Daily handwashing rates are calculated as the sum of all 'individual' uses over the course of each twenty-four hour period.

Alternative hygiene measures such as respondents' ratings of own handwashing habit formation, direct observation of respondent hand and nail cleanliness, and the presence of non-project liquid soap in the household were collected at the eight-month mark. We also collected measures of household sanitation, such as whether the household practices open defecation and whether they treat their water, to explore complementarities in behavior change and alternative mechanisms through which child health may be affected.

B. Child health

Incidence of child diarrhea and respiratory illness was collected at baseline and again every two weeks by surveyors, consisting of self-reports in which mothers were asked how many days each child had experienced diarrhea, loose stool motion, or the symptoms of respiratory illness in the past two weeks. These survey questions were adjusted at the eight month mark to account for many relevant cases being excluded given the strict initial definitions (described in detail in Section 6); this cross-sectional measure at eight months is our primary incidence outcome measure.

Anthropometric outcomes were collected at baseline and again at the eight month mark. These include child weight, height, and mid-arm circumference as measured by trained surveyors. We supplement self-reported incidence data with anthropometric outcomes to reduce the likelihood that any observed effects are driven by desirability bias on the part of mothers. Repeated diarrheal disease can affect child weight and height by reducing a child's ability to absorb sufficient nutrients from her food and thereby stunting her growth (McKay et al., 2010). We convert these measures into standardized height-for-age, weight-for-age, and midarm-circumference-for-age Z-scores (HAZ, WAZ, and MAZ, respectively) using the methodology provided in the WHO anthropometric guidelines; these Z-scores are calculated (as per WHO methodology) only for children ages 60 months and below (WHO, 2006).

B. Temporality of outcomes

Because various interventions were phased in and out at various times, below we define the time period for each effect of interest.

Baseline period is defined through the baseline survey, which was conducted four months prior to rollout.

Pre-change (rational habit formation) period is defined as the time between dispenser distribution and the monitoring service introduction/price change. We also zoom in on the three week period just prior to the date of change. This is because (1) we showed a video three weeks prior to the date of change to all dispenser-receiving households in order to increase and standardize comprehension regarding which treatment group each household was in; and (2) any rational habit formation effect should increase as the date of the anticipated change approaches.

Intervention period is defined as the four month period following rollout for IV1 (single ticket incentive) households and the four month period following the introduction of the monitoring service for MV2 (monitoring) households. For IV2 (triple ticket incentive) households

error in our feedback: we preferred that households were overcompensated for washing than undercompensated due to stricter and less verifiable measures of success (such as "all family members must wash", which is both harder to achieve and more difficult to verify), which in turn might diminish treatment effects.

only, this period is defined as the two months following the price change. These are the intervals within which pure intervention effects can be measured.

Persistence (habit formation) period is defined in the short run and the long run. The short run examines handwashing behavior within the month after the incentives or monitoring service are withdrawn; this is chosen to mirror the focused period of rational habit formation prior to the price change or commencement of the monitoring service. The long run examines handwashing behavior in the final month of data collection, which is nine months after the incentives are withdrawn (among IV1 and IV2 households) and seven months after the monitoring service is withdrawn (among MV2 households).

From the start of rollout to the end of the experiment, we accumulate a total of thirteen months of household-level data on handwashing behavior.

C. *Empirical strategy*

Our preferred specification for our behavioral outcomes, which primarily uses the dispenser sensor data, is as follows:

$$Wash_{hvt} = \alpha + \beta Treatment_{hv} + \gamma_t + \theta_v + \epsilon_{hvt} \tag{6}$$

in which $Wash_{hvt}$ represents the outcomes specified above, $Treatment_{hv}$ is the assigned treatment for each subset of comparisons described in Section 3, γ_t is day fixed effects, and θ_v is village fixed effects. The latter is included in all but those regressions comparing treatments across Monitoring and Incentive Villages (since randomization to MV or IV was at the village level). For analyses utilizing the midline survey, which is cross-sectional data collected eight months after rollout, we omit day fixed effects. Standard errors are clustered at the household level except in cross-IV-MV comparisons, in which they are clustered at the village level.

Our preferred specification for our child health outcomes, which primarily uses the midline cross-sectional data, is as follows:

$$Health_{cvt} = \alpha + \beta Treatment_{cv} + \delta BaselineHealth_{cv} + \theta_v + \epsilon_{hvt} \tag{7}$$

in which $Health_{cvt}$ represents the outcomes specified above (time varying for the biweekly measures, but cross-sectional for the single midline measure), $Treatment_{cv}$ is the assigned treatment group specified in the analysis, $BaselineHealth_{hv}$ represents the baseline value of the outcome variable, and θ_v is village fixed effects. Standard errors are clustered at the household level.

IV. Behavioral results

Table 2 presents a comparison of means between treatment and control households for baseline characteristics at the household, mother, and child levels. Treatment households are the pooled sample of all households who received the dispenser and soap. Appendix Tables 1A and 1B present the relevant pairwise comparisons for each sub-treatment arm. Households are balanced across most observables. Treated respondents are 0.4 minutes farther from their drinking water source, 3 percentage points less likely to be Hindu, marry 0.2 years later, rate themselves higher on whether people listen to them but lower on whether they make their children's health decisions, have taken their child to the doctor for an illness in the last two weeks 0.14 times more, and are 3 percentage points and 1 percentage point more likely to have a child experience a cold or diarrhea in the last two weeks, respectively. While the imbalance on the last three child health metrics may be concerning, the difference points in the opposite direction of the effects of interest, and we control for baseline health incidence in all health regressions. The disaggregated comparisons of Appendix Table 1A and 1B reveal no systematic differences across arms.

We next present our main results on the impact of each treatment on handwashing behavior. The description of time in all figures and tables henceforth will be relative to the date of the incentive price change or introduction of the monitoring service, denoted as Day 0. This aligns our experiment with the standard field experiment that begins when the intervention commences. In our setting, we begin our experiment 70 days *before* the key interventions of interest are implemented in order to explore whether agents are forward looking about future behavior change.

A. *Main treatment effects*

Incentives. — Table **??** presents results on the impact of the *extensive* incentives margin on handwashing behavior by comparing households in IV1.2, who received one ticket per day they washed at dinnertime (beginning on the day of rollout), with households in MV1.2, who received only the dispenser. Columns 1-4 of Panel A demonstrate that incentives worked as intended: after two months of incentives, incentivized households use the dispenser 1.7 more times over the course of the day than control dispenser households (Column 1), but this increase is not born out during the daytime (Column 2); rather, the bulk of the change in handwashing occurs around dinnertime (Columns 3 and 4). A similar pattern holds after four months of incentives (Panel B).

Appendix Figure A4 plots the raw time trend of handwashing during the daytime and the evening, respectively, across incentivized and dispenser only households over the four months that households were offered the one daily ticket incentive. While the response to incentives increases evening handwashing by approximately one press more per day relative to the control counterparts, there is no trend in daytime handwashing. Households appear to regard each handwashing event as an independent act. Though only suggestive, this underscores the importance of defining habitual behaviors with precision in behavioral change campaigns: to "wash hands before dinnertime" is perhaps a more tangible and trigger-centric instruction than the more popular instruction to "wash hands before eating, before cooking, and after defecation."

Column 4 of Table **??** use the preferred binary outcome variable of whether or not the dispenser was active during the household's stated dinnertime. Results show that incentivized households are 23 to 26 percentage points more likely than control households to wash at least once during their reported dinnertime (after two and four months, respectively). By the fourth month of incentives, just before the withdrawal of the intervention, incentivized households are washing their hands during their reported dinnertime 62% of the time.

Figure 2 plots the time trend of binary dinnertime handwashing rates across incentive and dispenser only households. The dashed vertical lines represent the average dates of surveyor visits, during which incentive households received calendar markings and tickets based on performance from the last batch of data collected. The time trend tells an important story. Households were first visited on Day -70: dispensers were delivered and incentive households were told about their daily ticket rewards, which they would begin earning immediately. They were next visited on Day -54, during which surveyors collected the first batch of handwashing data form the dispensers. On the third visit on Day -38, surveyors returned with the results of the first batch of data and the tickets the household had earned from this batch. Only upon *receiving* these tickets did households react to the incentive treatment. The reaction is followed by a steep decay, which is again buoyed by the next round of surveyor visits and tickets. Each of the third, fourth, and fifth visits prompt a sharp rise in handwashing, followed by an increasingly shallower decay. By the sixth round, despite continuing surveyor visits, household performance stabilizes.

This pattern is consistent with two stories. First, households may be building trust in the intervention. This is likely at the third visit but unlikely by the fifth. A complementary explanation is that surveyor visits serve as reminders or motivation to engage in handwashing. Motivation is particularly useful (as measured by the response to the visits) when the stock of handwashing that a household has accumulated is low in the early rounds. However, it becomes progressively less effective as the stock builds and the behavior becomes habitual. This pattern of "action and backsliding" is replicated in Allcott and Rogers (2014) in the tracking of household energy usage against the date of letters sent regarding energy consumption; the results are also consistent with a key prediction of Taubinsky (2014)'s model of inattentive choice and the substitutability of reminders and habituation.

We next move to the study of rational habit formation. In order to measure rational habit formation, we must first empirically establish two features of handwashing. The first is that handwashing can be moved by our chosen interventions of monitoring and tripling of incentives. If agents do not respond to these interventions, then we have failed to change the value of the behavior and agents have no reason to respond in anticipation. The second feature is that handwashing must be a behavior that can become habitual. If there exist no intertemporal complementarities in the behavior (measured by persistence after the withdrawal of interventions), agents gain no utility from accumulating handwashing stock prior to the introduction of the interventions.

We thus present our results from the intervention period first, then the persistence period, then return to the pre-intervention period to examine evidence of rational habit formation.

Intensive margin incentives. — We first examine the contemporaneous impact of an intensive-margin shift in incentives on handwashing. Columns 1 and 2 of Table 3 present the results for the comparison between households who were surprised with a triple ticket boost in incentives and those who remained with the single ticket incentive at Day 0. We report results both for the full 60 days during which households were exposed to the boost (i.e. earning triple tickets), as well as a lagged time frame of Days 30 to 59. The latter is relevant because Day 30 is the first day in which households who were eligible for tripled tickets on Day 0 physically received them. Households respond modestly to the tripling of daily tickets: they wash an average of 3 percentage points more than their single ticket counterparts over the duration of the triple ticket regime, increasing to a noisy 5 percentage points (8.5%) upon receiving the extra tickets in hand.

Figure 3 plots the three-day moving average of dinnertime handwashing rates for the tripled incentive arm relative to the standard incentive arm before and after the incentive boost. Note that the regression results of Table 3 control for average differences prior to the price change evident in the plot.

Monitoring. — Columns 3 - 5 of Table 3 estimate the contemporaneous impact of the monitoring service on household handwashing behavior as compared to dispenser-only households.[16] Column 3 presents results for the full tenure of the monitoring service, while Column 4 presents the lagged results of Days 30 to 59 and Column 5 presents results for the duration after calendar receipt. The monitoring service has a statistically significant and substantial impact on behavior, increasing handwashing rates by 8.5 percentage points (25%) in the comparable 30 - 59 day period.

Figure 4 presents the three-day moving average of dinnertime handwashing rates for monitored households relative to those who received the dispenser only. The graph demonstrates how household behavior to the monitoring arm reacts most strongly on the day of the first calendar receipt (Day 30), highlighting the role of a feedback mechanism in the effectiveness of the monitoring service. As with incentives, the regression results in Table 3 control for average differences prior to the intervention evident in the plot.

[16]Recall that the monitoring service lasted from Day 0 to Day 116, which is two months longer than the length of the triple incentive boost in incentive villages. This was implemented to compensate for the two months of incentives that all incentive households had already received prior to the boost, permitting a closer comparison between the long run effectiveness of incentives relative to monitoring.

B. Persistence

Section 5.1 establishes that the experiment exogenously increased the value and consequently the 'consumption stock' of handwashing in each treatment arm, albeit substantially more under the monitoring regime than the triple ticket regime. This addresses our first and second testable implications: $\frac{\partial d_t}{\partial x_t} \leq 0$. We now explore whether this exogenous shift in stock had an impact on subsequent handwashing behavior after the interventions ceased.

Many studies have examined the role of temporary interventions on persistence of behavior change (Charness and Gneezy, 2009; Conley and Udry, 2010; Dupas, 2014; Allcott and Rogers, 2014; Royer, Stehr and Sydnor, 2015; Aggarwal, Dizon-Ross and Zucker, 2020). The persistence of temporary interventions does not readily imply habit formation, however. Persistence can be generated by the purchase of a technology that changes the production function, the process of learning more about a technology (how to use it, what the optimal set of inputs is, or what the returns are) such that one updates her desire to engage, or the accumulation of consumption stock. Habit formation is driven only by the latter. Isolating this mechanism is a challenge, and existing studies lack the data or the context to distinguish the effects of consumption stock accumulation from learning or technology acquisition.

In the present study, we can easily rule out the first alternative mechanism behind persistence: because our outcome measure is the likelihood of dispenser use, it will not capture the effects of any other hygiene-related technology the household may acquire. Additionally, we find no changes in sanitation or water treatment practices (see Appendix Section 8.3), suggesting that households do not invest in alternative technologies that may alter their hand hygiene production function. In contrast, the mechanism of learning is a greater challenge to address, since the process of engaging in an activity repeatedly generates both learning about the activity and a growing stock of consumption.

We identify three dimensions of learning that can occur in our context: (1) learning how to physically wash one's hands, (2) learning how to use the handsoap dispenser, and (3) learning about the health returns to handwashing. We argue that the extent of learning required for the washing process is negligible: 99% of households already rinse their hands with water before mealtime, and 100% of households own and thus are familiar with the use of soap; to combine the two activities should require minimal learning and there is little reason to expect this to be differential across treatment groups. The extent of learning required to use the handsoap dispenser, which is a novel technology, may be greater; to address this, we allow a two-week learning period between the roll-out of the dispensers and the assignment to treatment during which all households can become acquainted with the dispenser. As is evident in Figure 2, households do indeed experiment with the dispenser technology for the first ten days, but behavior stabilizes thereafter, suggesting that this learning is largely complete within the first two weeks prior to treatment assignment.[17] Finally, households may persist in their handwashing because, by washing more, they also learn that handwashing leads to improvements in health, and therefore update their beliefs on the returns to the behavior. In Section 8.1 of the Appendix, we offer evidence that households who experience larger child health returns are no more likely to persist in their handwashing behavior than those who experience small child health returns, suggesting that this dimension of learning plays a minimal, if any, role in the persistence of handwashing behavior.

We therefore interpret persistence in handwashing behavior after the withdrawal of the interventions as evidence of habit formation: because the interventions that increased consumption stock are no longer active in this later time frame, any difference in performance between a treatment

[17]The same features of the behavior apply to handwashing as an experience good: perhaps households learn that soap does not ruin the taste of food or dry out their skin as much as they initially feared. The latter is unlikely because households regularly use soap in other contexts; the former is unlikely because households are familiar with non-scented soap (in fact, they request it in pilots) that would quickly assuage concerns around taste.

household and its relevant control must be due to intertemporal complementarities in the marginal utility of handwashing.

Table 4 presents the results on persistence. Results are separated into the first month after intervention withdrawal (short run: Panel A) and the final (thirteenth) month of the experiment (long run: Panel B). Column 1 of Panel A demonstrates that households who previously received the standard incentive continue to wash their hands during dinnertime 24 percentage points (63%) more than their dispenser-only counterparts during the first month after incentive withdrawal. The intensive margin of incentives, on the other hand, has no lasting effect as evident in Column 2 of Panel A: formerly triple-ticketed households continue to wash their hands slightly more (3.8 percentage points) than their single-ticketed counterparts in the month after withdrawal, but this is statistically indistinguishable from zero. Column 3 of Panel A demonstrates that, like the incentives on the extensive margin, the stock built from the monitoring intervention also persists: households are 9.3 percentage points (35%) more likely to wash than their dispenser control counterparts in the first month after the monitoring service is halted. These results confirm our third testable implication: $\frac{\partial d_t}{\partial k_t} \geq 0$.

In the long run, the magnitude of all effects decline. For both the incentives on the intensive margin and the monitoring service, we see no evidence of persistence nine and seven months (respectively) after intervention withdrawal (Columns 2 and 3 of Panel B). However, the effect of incentives relative to the dispenser control remains: households who had received the standard incentive nine months prior continue to wash 16 percentage points (121%) more than their dispenser control counterparts.[18]

These snapshots in time mask important trends in handwashing behavior over the seven to nine months of persistence observation. Appendix Figures A6 and A7 present the five-day moving average results for [formerly] incentivized and monitored households, respectively, relative to the dispenser control. Appendix Figure A8 presents the five-day moving average for all treatment arms, with incentive groups (IV1, IV2, IV2b) pooled and monitoring groups (MV2, MV2b) pooled. We highlight two observations. First, the dispenser-only households experience a secular decline in handwashing rates that parallels the decay of the other treatment arms. We suspect that this is due to an artifact of the experiment: after the withdrawal of the monitoring intervention, we reduced the frequency of our household visits (for all households in our sample) from twice monthly to once monthly. We sought to be as uninvasive as possible during the persistence period while still preserving access to the data. A consequence of this may have been a speeding up of the decay in handwashing across all arms. Second, the behavior of both formerly incentivized and formerly monitored households eventually converges to that of dispenser control households (approximately 30 days after intervention withdrawal for monitoring households and 100 days after for incentivized households). While they remain converged for formerly monitored households, formerly incentivized households more than double their handwashing rates relative to dispenser control households by Day 230, which corresponds neatly with the arrival of the winter season in West Bengal. It is likely that the uptick in handwashing reflects a habit that was attached not only to a time of day but also to a time of year and the associated contextual cues: cold weather, shorter days, and a visibly higher incidence of ARI symptoms (coughs, colds, and runny noses) in oneself and surrounding children. Though speculative only, this observation underscores both the value of long term, high

[18]A note on attrition: our sample size declined rapidly over the last several months as sensors malfunctioned. By the final month of data collection, we were left with 255 functional sensors, relative to an initial count of 1140. Appendix Table 2 explores the relationship between missing data and household handwashing performance: Column 1 presents the difference in dinnertime performance over the first nine months of the experiment between households that remained in the sample in the final month and those that did not (day fixed effects included); households that remained in the sample perform no better on average than their attrited counterparts. Columns 2 and 3 explore whether treatment effects are differential across attrited and remaining households; both coefficients are small, positive, and noisy, suggesting that households remaining in the sample experience treatment effects which are no larger than those that attrited. Appendix Figure A5 shows the trend in sensor attrition, which is largely balanced across treatment arms.

frequency data collection around behavior change (without which we may have assumed permanent convergence) and the contextual nuances of habit formation.

C. Rational habit formation

Having established that handwashing is a habitual activity and that the interventions change, to varying degrees, the value of handwashing, we now turn to the question of whether agents are rational about the habit-forming nature of this behavior. Results are presented in Table 5. We first examine the pre-change period. Recall that during this period, no incentive households had received the tripled tickets and no monitoring households had received a monitoring service. Rather, a portion of them had been notified on Day -54 (two weeks after rollout) that they should expect such a change to take place at a future date as circled on their calendar (Day 0). We compare the behavior of these anticipating households to households who were not told to expect any change in the future. Results are presented both for the full period of anticipation (Day -54 to Day -1) as well as for the final three weeks before the date of change (Day -21 to Day -1).

Columns 1-2 present the results for households anticipating a future tripling of tickets. We see no evidence that anticipation of a future price change affects current handwashing behavior. In fact, the coefficient becomes smaller as the date nears the date of change. The lack of behavior change in anticipation of tripled incentives is consistent with their having only a modest contemporaneous effect and no persistent effect: people did not appear to find the increase in incentives meaningful.

Columns 3-4 present the results for households anticipating a monitoring service. Recall that we find strong evidence that monitoring had a contemporaneous and also persistent effect, preconditions for anticipatory behavior. Consistent with rational habit formation and in contrast to the incentives setting, households anticipating monitoring are 4.7 percentage points (20%) more likely to wash their hands during dinnertime than their unanticipating counterparts; this rises to an 7.2 percentage point (34%) difference in the final three weeks before the monitoring commences.

Figures 5 and 6 depict these patterns graphically. Figure 5 plots the five-day moving average of handwashing behavior between anticipating and nonanticipating triple-ticket households. Household behaviors across the two arms follow essentially identical patterns before the price change and over the course of the experiment. In contrast, Figure 6, which plots the same for anticipating and nonanticipating monitoring households, depicts substantial additional stock accumulated in anticipation of the monitoring change, which appears to persist to some degree over the course of the intervention and thereafter. However, while the anticipation effect is clear, the results on persistence of anticipatory stock are noisy and small, as demonstrated in Appendix Table 3.

Lastly, we elicited biweekly forecasts of handwashing to measure the degree of sophistication or intention-setting households may have regarding their future handwashing behavior. Results are broadly consistent with actual performance: anticipating households in the monitoring regime intend to wash more prior to the onset of monitoring, while anticipating households in the triple-ticket regime make no such intention. However, we interpret the forecasting data with considerable caution given the challenge of eliciting meaningful forecasts in the field. The results are discussed in detail in Appendix Section 8.2 and reported in Appendix Table 8.

D. Discussion

While our results across contemporaneous, persistent, and anticipatory behavior change are consistent with a story of rational habit formation, we explore three alternative mechanisms which may contribute to the patterns we observe in the monitoring treatment arm: confusion, salience, and reciprocity.

Confused households may have believed that they were being monitored starting on Day -54 rather than Day 0. The experiment embedded a series of measures to alleviate this concern: first, the future date of change was circled in red on the household calendar and written on a sticker attached to the dispenser on Day -54. Second, households were shown a video on Day -21 reiterating their treatment assignment; the videos involved comprehension questions where the respondent confirmed the date on which the enumerator would "turn the switch on" to begin monitoring or incentives would be tripled. Third, households were reminded of their treatment assignment on every surveyor return date (Day -54, Day -38, and Day -21): as an anticipating household, one was reminded of the upcoming date of change; as a non-anticipating household, one was reminded that the surveyor would continue to return every two weeks to collect child health data. Finally, if households did indeed believe that the monitoring service started on the day of announcement rather than Day 0, then we should expect their patterns of response to be qualitatively similar to those of households who actually did receive the monitoring (with no anticipation) on Day 0. However, if we compare the behavior of MV3.1 to that of MV2.2 (see Figure 6), we see little in common: anticipating households respond sharply on visit days and decay nearly as sharply afterwards, while households actually being monitored respond sharply on visit days and remain responsive, steadily increasing their handwashing rates over time. Although only suggestive, the two groups are randomly assigned and the difference in patterns is stark and difficult to reconcile with a story of confusion among anticipating households.[19]

An alternative mechanism for the apparent anticipatory reaction among monitoring households is differential salience. Perhaps the term "future monitoring" made the act of handwashing more salient today than the term "future tripling of tickets" or the absence of language around interventions. Similarly, perhaps enumerators conveying the future monitoring intervention placed more weight on present handwashing.[20] Salience alone plainly impacts handwashing behavior as evidenced by the usage spikes on the day of visit (recall the patterns of Figure 2 among extensive incentive households). Among anticipating households, usage spikes on the pre-change visit days can be interpreted in two ways (neither mutually exclusive). First, the presence of enumerators and the reminder of the impending future Day 0 may have brought the future to top-of-mind, a necessary ingredient to forward-thinking behavior in a world of limited cognitive capacity, and therefore consistent with rational habit formation. Second, enumerators may have differentially emphasized *contemporaneous* washing for those who were anticipating the impending Day 0 change, yielding patterns consistent with rational habit formation but in fact due to differential salience of contemporaneous washing behavior. On visit days across all treatment arms, surveyors asked households about how the dispenser was functioning and whether there were any problems; they asked households to forecast how many days mothers and children anticipated washing their hands before mealtime in the coming week; they then [openly] marked the level of soap on the tank to underscore the importance of soap use, replenished the soap, and performed "maintenance" on (i.e. collected data from) the dispenser. Given this standardized procedure across all households which emphasized the practice of handwashing, it is unlikely that enumerators placed additional and differential weight on contemporaneous washing among anticipating monitoring households only (relative both to nonanticipating households and to triple ticket anticipating households). However, without voice recordings of enumerator behavior for each visit, we are unable to decisively rule out this mechanism. While we took all possible steps to maintain equivalent salience across

[19]Alternatively, one may think that households were not confused, but simply did not trust us and therefore believed we were already monitoring them. If this were the case, then come day -54 or -21, when households faced no repercussions to being monitored, no feedback, and a continual insistence that a monitoring service would be provided on Day 0, suspicion should have decreased over time. In contrast, we see a steady rise in responsiveness to the anticipation over time relative to the non-anticipating group.

[20]The challenge of salience plagues not only most experimental literature in which individuals are responding to various types of information but also all of the empirical rational addiction literature, where announced future price changes may make the current activity engaged in more top-of-mind. Existing literature in rational addiction has the compounded salience confound that an exogenous price change may be interpreted as a signal of how valuable or harmful a good is, causing agents to update their cost function and react accordingly. Our experimental design minimizes this confound by making the treatment assignment an explicit lottery.

treatment arms, it is possible that differential salience of contemporaneous handwashing is an alternative channel driving the spikes in the anticipated monitoring group prior to the monitoring intervention.

Lastly, we consider the alternative mechanism of reciprocity: perhaps households who were anticipating a service from our enumerators wished to express their gratitude by raising their contemporaneous levels of handwashing. Households were likely to view the monitoring intervention as a service rather than a threat, so this mechanism is difficult to rule out entirely. One might expect, however, that such reciprocal behavior would be likewise activated among unanticipating households upon being told that they would receive the monitoring service immediately (on Day 0). Among these households, the monitoring intervention was explicitly marketed as a gift from the NGO; because we wished to help them wash their hands and were able to spare the additional time and effort, we wished to provide the service to them. This is distinct from the unanticipating households, who knew that their names were randomly drawn from a lottery for the service. We do not observe reciprocal behavior among surprised households, who only react once they receive a calendar in their hands thirty days later.[21]

While confusion, differential salience, or reciprocity may each play some part, various features of our design and multiple sources of evidence suggest that they are unlikely to be the driving force behind the results we observe. Our behavioral results instead appear to most closely align with a model of rational habit formation. While the contemporaneous effects of monitoring were substantial (8.5 pp, or 34% more handwashing than the relevant control mean), those of the tripling of tickets were smaller (5.1 pp, or 9% more handwashing than the relevant control mean). While the persistence of monitoring was clear, the tripling of tickets had no persistent effects. Consistent with the utility function of a rational habit former, households chose not to invest in accumulating handwashing stock to 'prepare' for an intervention with little contemporaneous benefit or prospects of habit formation. On the other hand, they invested considerably in accumulating stock for an intervention with significant contemporaneous and long run bite.

That households exhibit some detectable level of rational habit formation should be, upon reflection, not so surprising. First, our design sets up the optimal scenario to facilitate rational habit formation: households are fully aware that we want to help them develop a habit of handwashing, and we reiterate the future dates at which the value of the behavior will change. Second, though rational habit formation at first glance appears to require considerable sophistication, we manifest this behavior in many parts of our lives without articulating it as such: we train ourselves to eat less, wake up earlier, speak in a particular style, or sit in a particular manner when we know there is a high-stakes opportunity approaching in which to do so would reap rewards (or equivalently, save us from punishment). None of these behaviors are ones we must learn per se (we could, with enough focus, activate them on the spot), but we recognize that practice will make it easier to engage when the time comes. It appears that handwashing for a community in rural West Bengal likewise falls in this category: agents internalize, at least in part, the habitual nature of the behavior and accumulate stock accordingly.

V. Health results

Thus far, we have established that handwashing is a habit-forming behavior and that households in our sample are sophisticated about its nature. It remains an open question, however, whether the

[21]One tension in this set of results is that, while households who were expecting future monitoring reacted in anticipation, upon receiving the monitoring, households did not respond until the receipt of the calendars thirty days later. The execution of the monitoring intervention helps resolve this tension: most simply, the fear or anticipation of the event was likely stronger than the event itself in our setting. While households experienced three weeks of anticipating an intensive monitoring intervention, the actual commencement of the intervention on Day 0 was likely underwhelming, since the only action that occurred on Day 0 was that the switch was [supposedly] turned on. Not until Day 30 did all the anticipated dimensions of the monitoring service transpire.

habit of handwashing is worth acquiring in the high-disease environment of West Bengal, where the marginal value of this simple activity may be small given the high exposure to disease from other sources. We now ask: does handwashing generate positive health benefits, $\alpha > 0$? We examine three sets of data. The first utilizes day-level reports by mothers of child diarrhea and acute respiratory incidence (ARI) as collected by surveyors every two weeks during the first five months of the experiment. We examine health data from months four and five only, as this encompasses the peak of handwashing performance across treatment households.[22]

Our second set of outcomes utilizes two-week incidence reports from the midline survey conducted between months seven and eight. This midline survey revised the manner in which we collected data on child diarrhea and ARI. The restructuring was motivated by concerns from the field that surveyors were missing incidence cases. For example, for diarrhea, reporting mothers (1) felt diarrhea was a serious illness that their children could not suffer from unless the child was visibly sick and (2) often did not know whether their children had experienced regular loose stool motions since their children played outside most of the day and defecated in open fields away from the house. We therefore revised the questions to cast a wider net on illnesses and we required surveyors to have the child present during the time of surveying.[23]

Our third set of health outcomes is drawn from the midline survey as well: we recollect anthropometric measures of child height, weight, and mid-arm circumference.[24] We combine these measures with child age and gender data to create the child's height-for-age, weight-for-age, and mid-arm circumference-for-age Z-scores based on guidelines by the World Health Organization (WHO). This analysis is restricted to children 0 to 60 months as specified by the WHO guidelines.[25].

Appendix Table 4 presents the intent-to-treat estimates from the child-day level incidence reports. All regressions include day and village-level fixed effects as well as a full set of child health baseline controls, although results are robust to excluding baseline controls (not shown). Columns 1 and 3 report results for the pooled sample of all treated households relative to households in the pure control group respectively for diarrhea and ARI incidence. Columns 2 and 4 disaggregate this sample into each treatment group: incentives, monitoring, and dispenser control households.

While estimates for the impact of treatment on diarrhea are consistently negative, they are noisy and close to zero. This is not surprising, as the reported likelihood of a child in the pure control group suffering from diarrhea on a given day is only 0.4 percent. Results on ARI are clearer: children in treated households are 1.9 percentage points (13%) less likely to be suffering from ARI on a given day than their untreated counterparts; this effect size, significant at the one percent level, is relatively evenly distributed across the treatment groups, with monitoring households seeing the

[22]Although this time restriction was not specified in the pre-analysis plan, we did not explore any other time-frame for the health outcomes during our analysis to avoid the concern of multiple hypothesis testing.

[23]The wider net was cast as follows: mothers and children together were asked whether the child had experienced *any* loose stool motion in the last two weeks. If so, the days they experienced loose stool were recorded. This is in contrast to the previous five months of incidence data collection, during which mothers (and not children) were asked whether their child had experienced loose stool motion at least three times in a day, the clinical definition of diarrhea. Any amount less than three was not recorded as an episode. As is evident in Appendix Table 4, this yielded too few cases for statistically significant movement to be detectable. We acknowledge that a single loose stool motion is not necessarily reflective of diarrhea; however, a single reported motion is likely to be a signal of more actual motions in a day (given the recall problem for young children and the lack of supervision by mothers). We report the results, however, as 'loose stool' and not as 'diarrhea' and leave the reader to interpret. For the ARI question, mothers and children together were asked whether the child had experienced *any* of the symptoms of ARI in the last two weeks, and the surveyor listed the following: runny nose, nasal congestion, cough (with or without sputum production), ear discharge, hoarseness of voice, sore throat, difficulty breathing or a prescription from a doctor for such. If the respondent answered yes to any of these symptoms, the surveyor then asked how many days the child had experienced these symptoms. This is in contrast to the previous five months, during which surveyors asked whether the child had suffered from any two of the three symptoms of a runny nose, cough, or fever.

[24]We had an attrition rate of 9.6 percent in the midline survey. Appendix Table 5 demonstrates that the households surveyed at midline were no different from one another on baseline outcomes across treatment assignment.

[25]The WHO provides the distribution of each age and gender-specific anthropometric measure for a reference population of well-nourished children from Brazil, Ghana, India, Norway, Oman and the United States, such that a Z-score of 0 is the median of the reference population. We place a special focus on height-for-age as a metric of child health improvement, as linear growth is regarded as the most relevant indicator of overall nutrition (Hoddinott et al., 2013)

largest drop in ARI incidence of 2.9 percentage points, or 19%.

Table 6 presents the intent-to-treat estimates from our preferred restructured midline survey with and without baseline controls. Mean two-week incidence of loose stool in the pure control group is 9.9%. A child in a treated household is 2.5 percentage points (25%) less likely to experience loose stool motion in the previous two weeks. Similarly, the average treated child experiences 0.08 fewer days of loose stool (38%) per two weeks, significant at the five percent level. When we broaden the net to any loose stool, the impact of handwashing is clear.

ARI results remain consistent in percent magnitude with those in Appendix Table 4. A child in a treated household is 4.1 percentage points (23%) less likely to show any symptoms of ARI in the last two weeks and experiences 0.2 fewer days (16%) of ARI per two weeks. Appendix Table 6A disaggregates these results into each treatment arm; treatment effect sizes remain broadly consistent across arms, but we do not focus on these given concerns over multiple hypothesis testing.

Table 7 presents the intent-to-treat estimates on child anthropometric outcomes. Weight-for-age z-scores increase by 0.13 , height-for-age z-scores by 0.21, and mid-arm circumference-for-age z-scores by 0.06. To get a sense of the magnitude of these results, consider that children ages five years and below in treated households are approximately 0.38 kg heavier than those in pure control households. At a conversion rate of 7780 calories per kilogram (Wishnofsky, 1958) and given that the dispensers have been in use for eight months at the point of data collection, treated children are able to absorb approximately 12 more calories per day than children without a dispenser.[26] Appendix Table 6B disaggregates these results into each treatment arm, and Appendix Table 6B disaggregates by age of child. Unsurprisingly, younger children (one to two years of age) benefit most in weight, height, and mid-arm circumference.

Since the average rate of handwashing at dinnertime among treated households is 47%, we now consider estimates of the treatment on the treated (TOT). However, because control households were not given a dispenser, we cannot employ dispenser use as a proxy for handwashing in this instrumental variables exercise. Instead, we employ an alternative hand hygiene measure we collected across all sample households: self-reports on whether the mother and child wash regularly (whether they have achieved a handwashing habit). This measure is highly correlated with dispenser use (Appendix Table 9A, Column 3). For ease of interpretation, we transform the measure into a binary variable such that the self-report is equal to one when the respondent articulates that a habit has been achieved and zero otherwise. This binary washing variable serves as an instrument for whether or not a household received a dispenser.

In particular, we run the following two-stage regression for child c in household h, village v, and time t:

$$\begin{gathered} Wash_{hv} = \alpha + \beta_1 TreatedHousehold_{hv} + \epsilon_{hv} \\ Health_{chv} = \alpha + \beta_2 \hat{Wash}_{hv} + \delta_{chv} + \theta_v + \epsilon_{chv} \end{gathered}$$

in which $Wash_{hv}$ is the self-report, δ_{chv} is a vector of child health baseline controls and θ_v represents village fixed effects.

Table 8 presents the TOT estimates. A child in a household that reports regularly washing at dinnertime experiences a 48% decrease in the likelihood of having loose stool, a 72% decrease in the number of days she experiences loose stool, a 30% decrease in the likelihood of experiencing any ARI symptoms, and 30% fewer days of ARI. She also sees a 0.21 standard deviation rise in her WAZ score (noisy) and a 0.24 standard deviation rise in her HAZ score (noisy).

The preceding analysis yields two key takeaways. First, the results provide the first causal evidence in the literature that handwashing alone generates significant positive health benefits in the developing world. Second, the analysis highlights that the mere provision of the dispenser and

[26]This exercise was adopted from Bennett, Naqvi and Schmidt (2018), and despite significant differences in the type and time length of handwashing interventions being tested between this paper and Bennet et al., the estimated change in per day caloric intake due to the intervention is remarkably similar (12 v. 14 calories per day).

liquid soap has a significant impact on child health. In fact, the marginal impacts of each treatment arm are for the most part statistically indistinguishable from the impact of the dispenser arm alone. This large treatment effect to dispenser provision cannot be due to a newfound availability of soap in treated households, as baseline estimates point to 99% of households having [and using] soap in the home. Rather, this must be due to some combination of the household's valuation of the dispenser and liquid soap and thereby the act of handwashing ("if we receive something so nice, handwashing must be important and we should use it") along with the convenience of the dispenser location, being stationed right next to the place of eating. Novelty is a less likely explanation, since our results are estimated seven to eight months after the distribution of the machines.

VI. Conclusion

This study analyzes the process of habit formation in the high-impact preventive health behavior of handwashing with soap, examining how individuals internalize and interact with this habit-forming behavior. Our results suggest that monetary incentives and monitoring and feedback are effective means of increasing handwashing rates in the short run. While the impact of incentives on the extensive margin is substantial, intensive-margin changes in incentives have diminishing returns. Both extensive-margin incentives and monitoring have persistent effects, establishing that handwashing is indeed a habitual behavior. We also present evidence that agents internalize the intertemporal complementarities in the marginal utility of handwashing. Specifically, households respond strongly in anticipation of a future monitoring intervention, but show no response in anticipation of a future intensive-margin change in incentives. This is consistent with the theory of rational habit formation, in which agents should only respond in anticipation to interventions that alter the consumption value of future behavior and exhibit habit-formation through persistence.

This exercise offers the first well-identified estimate of the presence of rational habit formation, and additionally for *good* habits, in the literature. These findings inform the optimal incentive design of programs that seek to increase the takeup of good habits. Namely, if a behavior is habit-forming, then an intervention may do better to front-load incentives, and thereby maximize habit stock, rather than spread incentives over time. If individuals are rational regarding the habitual nature of the behavior, incentives that are offered at a future date will generate a larger stock of consumption in the long run than those administered immediately. The optimal type, size, and length of such incentives remain important areas of future exploration.

This paper also sheds light on the production function for child health as it relates to the input of hand hygiene. We establish the strong link between handwashing with soap and child incidence of respiratory infection and diarrhea, and the experiment can uniquely offer treatment-on-the-treated estimates which suggest that a child who achieved a regular dinnertime handwashing practice saw a 72% decrease in the number of days she experienced loose stool motion and a 30% decrease in the number of days she experienced acute respiratory infection. These translate into substantial improvements in anthropometric measures that have long run implications for the health of the child: receiving a handsoap dispenser and liquid soap generates, within eight months, improvements in child mid-arm circumference-for-age z-scores of 0.06, child weight-for-age z-scores of 0.13, and child height-for-age z-scores of 0.21. These findings point to the importance of human-centric design: dispenser provision was not effective because it provided the households with soap; rather, the location, ease of use, and attractiveness of the dispenser and soap must have motivated the practice of handwashing. While this study was not designed to identify these effects, at the fixed cost of $4.00 USD per dispenser and variable cost of $1.00 USD per 15 liters of foaming soap per year (the average household consumption rate), such product-design mechanisms are additionally a fruitful avenue of future exploration.

VII. Figures

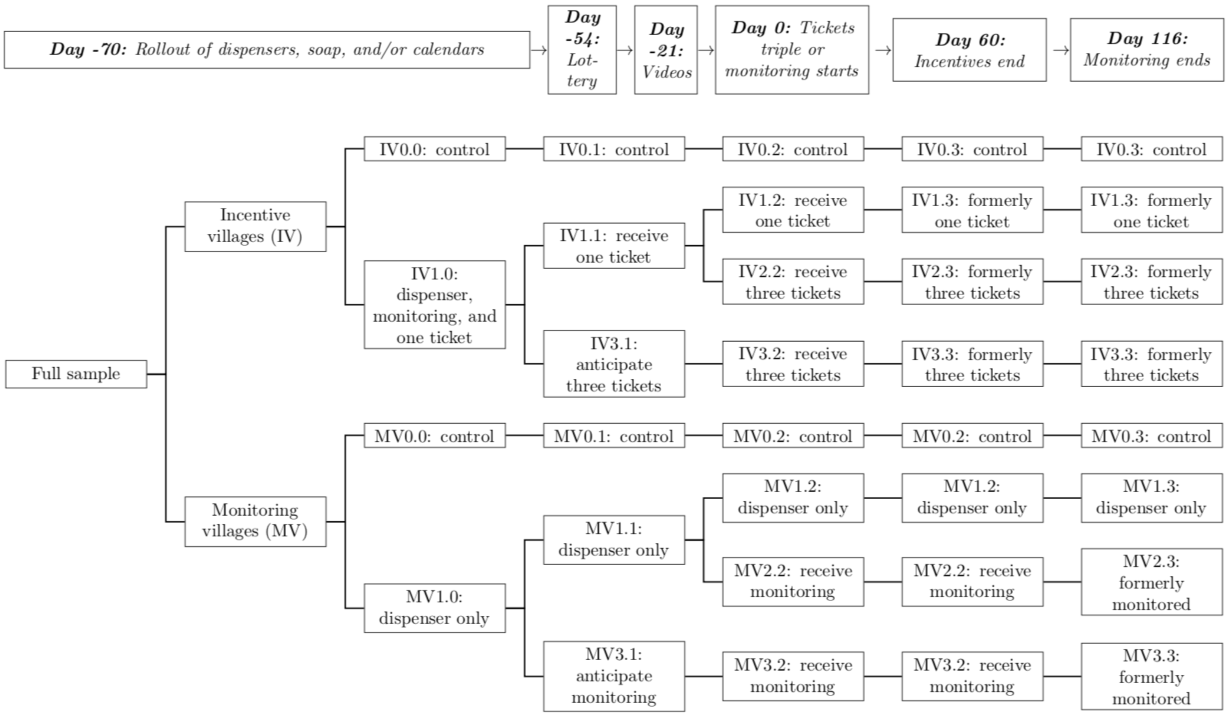

Figure 1. : Randomization map

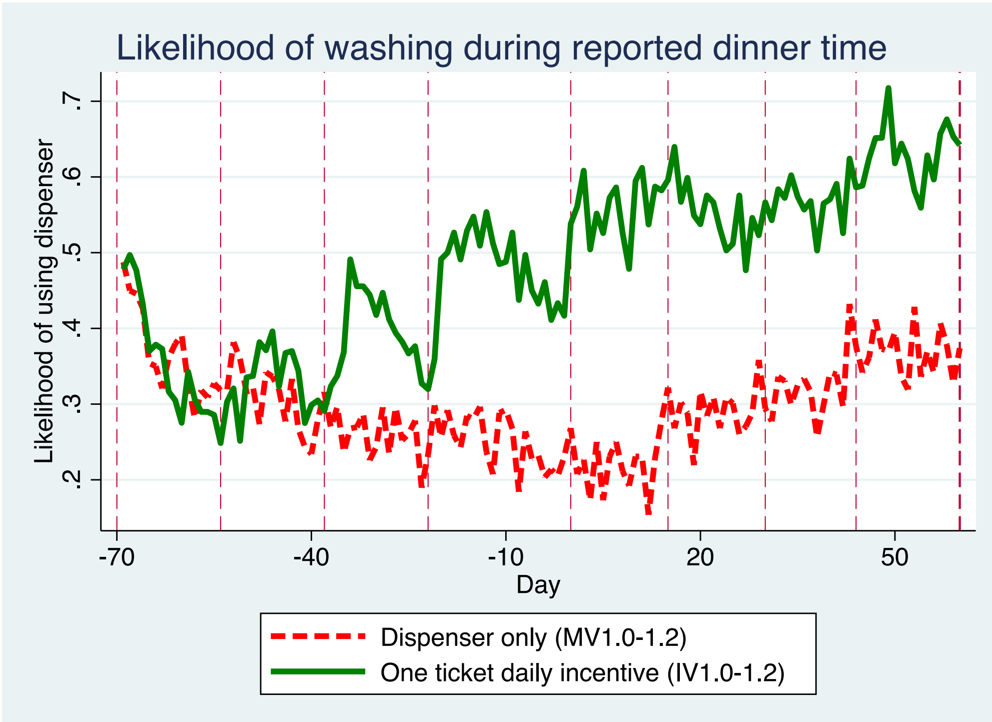

Figure 2. : Binary use at dinnertime

Note: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Red dashed line represents households who received only the dispenser; green line represents households who received the dispenser, feedback, and one ticket for every night the dispenser was active around their self-reported dinnertime. Vertical dashed represent the approximate surveyor visit day. Day 0 marks the day that incentives were tripled or monitoring commenced for a subset of households (neither subset is shown in this figure). Day -70 is the day of rollout. Tickets were distributed for the full length of the graph shown (until Day 60).

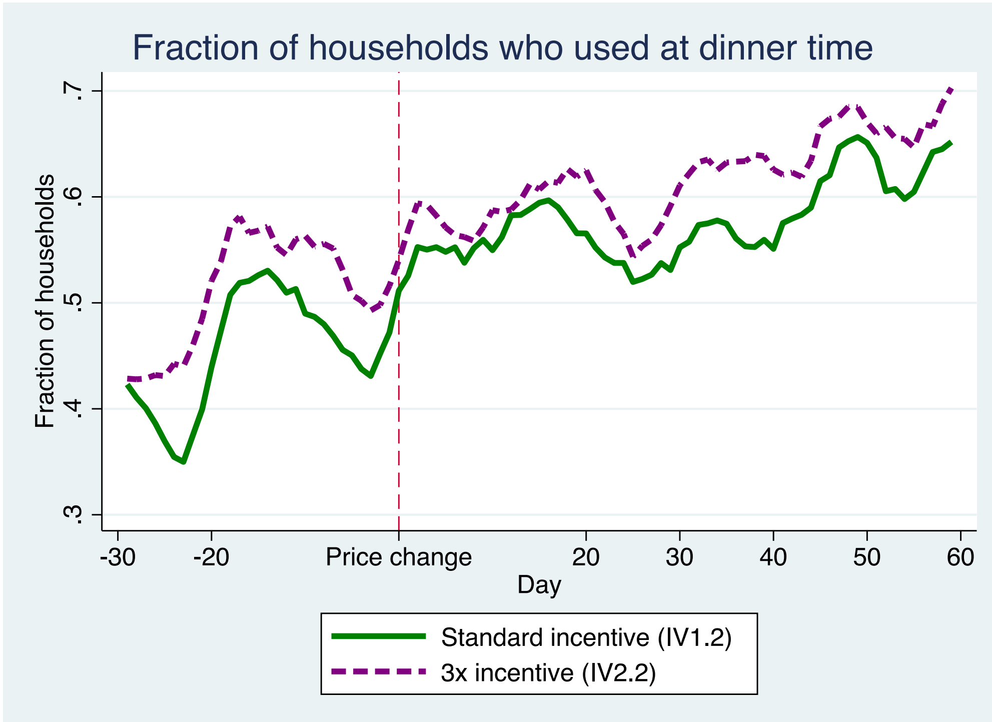

Figure 3. : Incentive effect during intervention regime

Note: Figure shows the five day moving average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Green line represents households who received the dispenser, feedback, and one ticket for every night the dispenser was active around their self-reported dinnertime; purple dashed line represents households who received one ticket until the point of the "Price change" (Day 0) and received three tickets for every night the dispenser was active during dinnertime for the remainder of the days displayed in the figure.

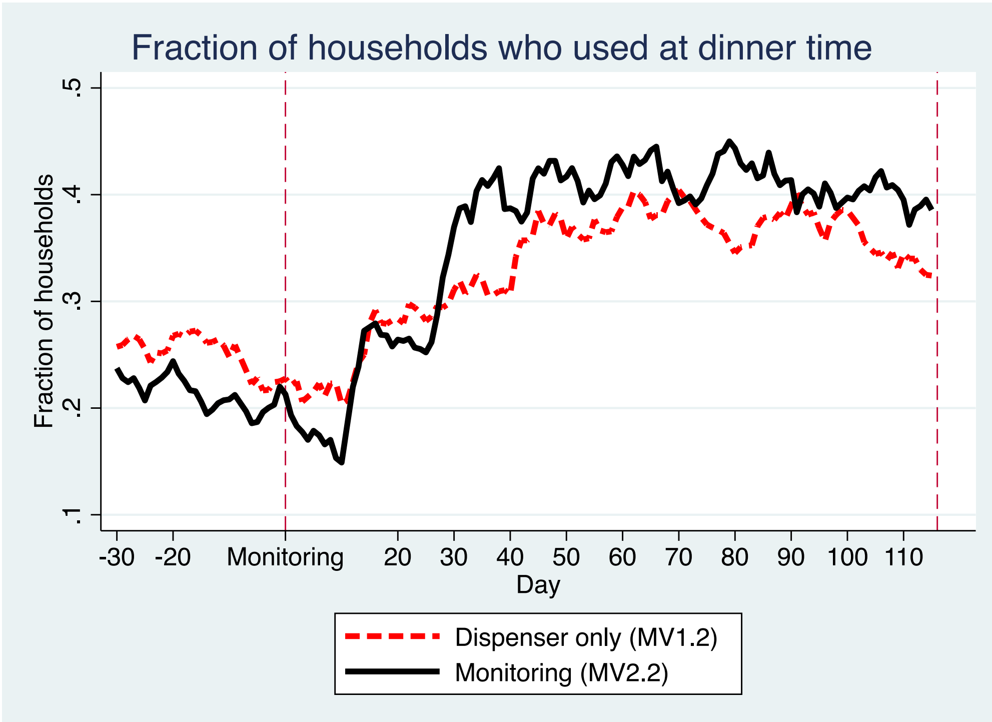

Figure 4. : Monitoring effect during intervention regime

Note: Figure shows the five day moving average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Red dashed line represents households who received the dispenser only; Black line represents households who received the dispenser only until the point of the "Monitoring" (Day 0) and received feedback/monitoring on behavior thereafter for the duration displayed in the figure.

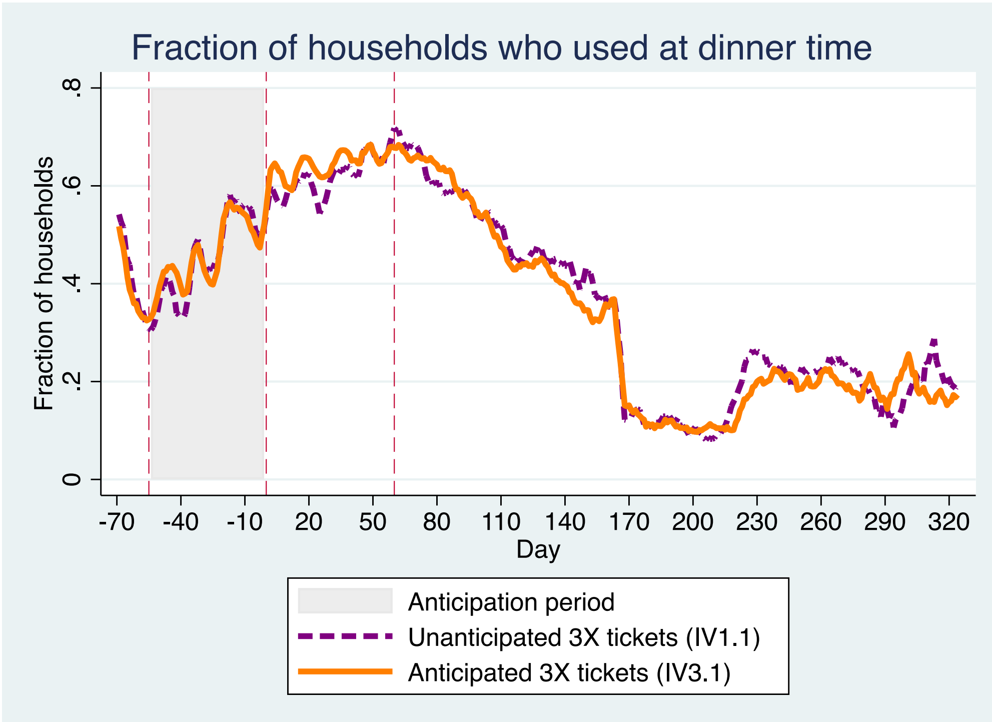

Figure 5. : Rational habit formation in incentives

Note: Figure shows the five day moving average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Both purple and orange lines represent households who received the dispenser, feedback, and one ticket until Day 0, after which they received three tickets per day the dispenser was active during the evening mealtime; however, orange households were anticipating the tripling of the tickets while dashed purple households were not. The gray box represents the time during which orange households were anticipating. Triple tickets then commenced on Day 0 and lasted until Day 60 (third vertical red line).

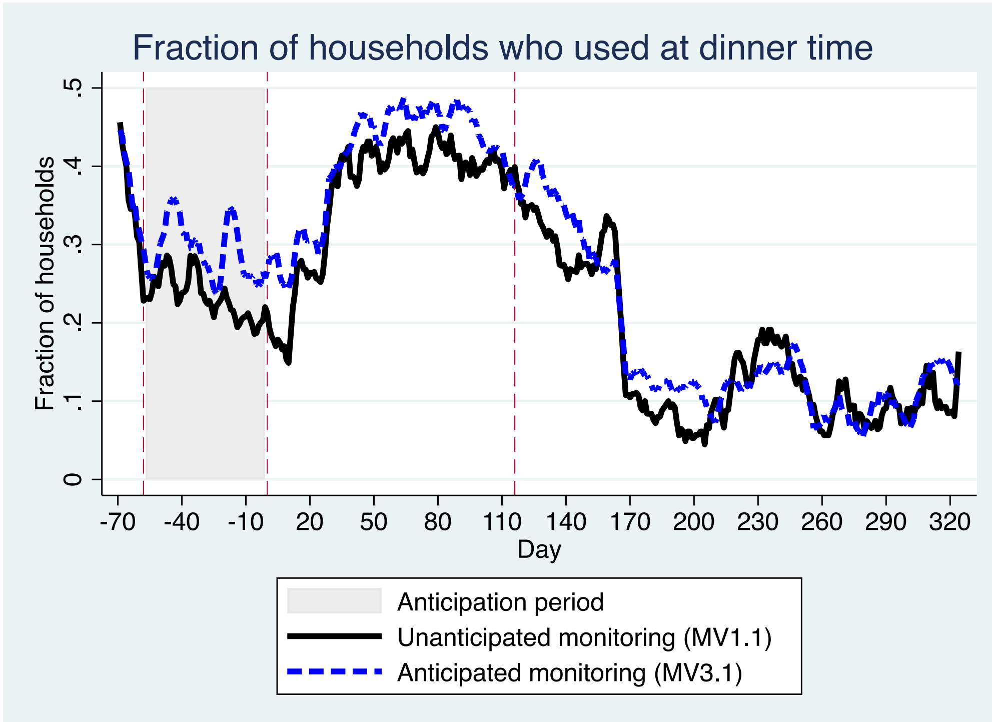

Figure 6. : Rational habit formation in monitoring

Note: Figure shows the average likelihood of the dispenser being active (at least one press) 1.5 hours before or after the household's self-reported evening mealtime. Both black and blue lines represent households who received the dispenser only until Day 0, after which they additionally received feedback/monitoring; however, dashed blue households were anticipating the start of monitoring/feedback while black households were not. The gray box represents the time during which dashed blue households were anticipating. Feedback then commenced on Day 0 and lasted until Day 117 (third vertical red line).

VIII. Tables

Table 1—: Balance across treated and control

	(1) Pure control mean	(2) Treated mean	(3) p-value	(4) N
Panel A: Household				
Access to electricity	0.9537	0.9569	0.9337	2885
Daily labor occupation	0.5429	0.5479	0.8499	2887
Agriculture occupation	0.2177	0.2076	0.7513	2887
Number of rooms	2.0641	2.0748	0.2517	2884
Deep tubewell drinking source	0.5583	0.5620	0.7826	2887
Distance to drinking source (min)	9.2632	9.6866	0.2272	2884
Latrine	0.3785	0.3738	0.3182	2886
Mobile	0.7697	0.7581	0.4672	2887
Breakfast start hour	8.0278	8.0727	0.0316	2881
Lunch start hour	12.9181	12.9564	0.3203	2881
Dinner start hour	20.3666	20.3804	0.5266	2888
Panel B: Hygiene and sanitation				
Cold can spread	0.6117	0.6056	0.9720	2887
Soap cleans germs from hands	0.9457	0.9455	0.7518	2888
Number of times hands washed	2.7013	2.6887	0.4280	2888
Open defecation practiced	0.6836	0.6761	0.9815	2887
Panel C: Mother				
Age (years)	31.6356	31.8338	0.3439	2888
Education (years completed)	6.0246	6.0290	0.5978	2886
Hindu	0.7273	0.6992	0.6376	2886
General caste	0.3356	0.3539	0.6124	2882
Age at marriage	16.4111	16.6295	0.0653	2868
People listen	2.9989	3.0546	0.1395	2886
Mother makes child health decision	3.3537	3.1912	0.0811	2882
Panel D: Children below 11 years				
Age of child (months)	69.4566	69.4133	0.9526	4797
Male child	0.5000	0.4939	0.8162	4803
Height (cm)	104.7536	105.2649	0.2166	4789
Weight (kg)	15.2206	15.1965	0.9691	4788
Preventive check-up (no. of times 6 mo.)	0.7543	0.7020	0.7267	1736
Sick doctor visit (no. of times 6 mo.)	1.6607	1.7925	0.2447	1694
Had cold in the last two weeks	0.3550	0.3863	0.0857	4795
Had cough in the last two weeks	0.0759	0.0847	0.2558	4739
Had diarrhea in last two weeks	0.0476	0.0586	0.2503	4800
Exclusively breastfed (no. of months)	4.6940	4.6090	0.5278	3195

Notes: Treated" pools all households that received a dispenser. "Pure control" are households who did not receive a dispenser. p-values computed in a regression of the variable on treatment assignment with village level fixed effects.

Table 2—: Impact of incentives on the extensive margin

	(1) Total Daily Presses	(2) Total Presses before 5	(3) Total Presses after 5 pm	(4) Likelihood of use during reported dinnertime
Panel A: Two-month mark (Day -10 to 0)				
One ticket daily incentive (IV1.2)	1.6836***	0.3956	1.2879***	0.2275***
	(0.611)	(0.483)	(0.213)	(0.035)
Mean of pure control				
	3.938	3.139	0.799	0.235
	[6.275]	[5.780]	[1.807]	[0.424]
N				
N2	3389	3389	3389	3186
Panel B: Four-month mark (Day 50 to 59)				
One ticket daily incentive (IV1.2)	1.2314**	-0.0241	1.2555***	0.2599***
	(0.668)	(0.512)	(0.260)	(0.044)
Mean of pure control	6.180	4.525	1.655	0.364
	[8.626]	[7.384]	[3.423]	[0.481]
N	3165	3165	3165	2955

Notes: Observations are at the household-day level. Robust standard errors in parentheses and clustered at the village level. Standard deviation in brackets. All regressions include fixed effects for day. All regressions include fixed effects for day. Households in the one ticket daily incentive group are compared to households in the dispenser only group. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). One MHT family: the 8 coefficients of the effect of the IV1.2 treatment on handwashing outcomes (Columns 1-4 of Panels A and B). * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 3—: Impact of incentives on the intensive margin and monitoring service

	(1) Day 0 to 59	(2) Day 30 to 59	(3) Day 0 to 59	(4) Day 30 to 59	(5) Day 30 to 116
	Likelihood of using during reported dinnertime				
Contemporaneous tripled incentive (IV2.2)	0.0273	0.0509			
	(0.023)	(0.026)			
Contemporaneous monitoring (MV2.2)			0.0548***	0.0846***	0.0844***
			(0.022)	(0.027)	(0.025)
Mean of pure control	0.578	0.598	0.312	0.346	0.361
	[0.494]	[0.490]	[0.463]	[0.476]	[0.480]
Comparison group					
N	18542	9901	11597	6563	19098

Notes: Observations are at the household-day level. Robust standard errors in parentheses are clustered at the household level. Standard deviation in brackets. All regressions include village and day fixed effects. All regressions control for average dinnertime handwashing rates prior to price boost or commencement of the service, which occurred on Day 0. Control group for columns 1-2 is the standard (1 ticket) incentive treatment arm. Control group for columns 3-4 is the dispenser only arm. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Two MHT families: the 2 coefficients of the effect of the IV2.2 treatment on handwashing outcomes (Columns 1-2); and the 3 coefficients of the effect of the MV2.2 treatment on handwashing outcomes (Columns 3-5). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 4—: Persistence in handwashing after withdrawal of interventions

	(1)	(2)	(3)
	Likelihood of using during reported dinnertime		
	Day 60 to 89	Day 60 to 89	Day 116 to 146
Panel A: Short run			
Former single ticket incentive (IV1.3)	0.2403***		
	(0.036)		
Former tripled incentive (IV2.3 + IV3.3)		0.0387	
		(0.025)	
Former monitoring (MV2.3 + MV3.3)			0.0928***
			(0.026)
Mean of pure control	0.379	0.619	0.269
	[0.485]	[0.486]	[0.443]
Comparison group			
N	8589	18439	10837
	Day 300 to 329	Day 300 to 329	Day 300 to 329
Panel B: Long run			
Former single ticket incentive (IV1.3)	0.1609**		
	(0.073)		
Former tripled incentive (IV2.3 + IV3.3)		-0.0539	
		(0.064)	
Former monitoring (MV2.3 + MV3.3)			-0.0371
			(0.029)
Mean of pure control	0.133	0.295	0.133
	[0.340]	[0.456]	[0.340]
Comparison group			
N	1726	3390	3080

Notes: Observations are at the household-day level. Robust standard errors in parentheses and clustered at the village level for Column 1 and at the household level for Columns 2-3. Standard deviation in brackets. All regressions include day level fixed effects; columns 2-3 also include village level fixed effects. Comparison group for "Former single ticket incentive" and "Former monitoring" is the dispenser only group; comparison group for "Former tripled incentive" is the former single ticket incentive. Panel A estimates effects during the first month after the withdrawal of the relevant intervention; Panel B estimates effects from the final month of data collection (which is nine months after the withdrawal of incentives and seven months after the withdrawal of monitoring). p-values adjusted for multiple hypothesis testing (MHT) using Anderson (2008). Three MHT families: the 2 coefficients of the effect of the IV1.3 treatment on handwashing outcomes (Column 1 of Panel A and B); the 2 coefficients of the effect of the IV2.3+IV3.3 treatment on handwashing outcomes (Column 2 of Panel A and B); and the 2 coefficients of the effect of the MV2.3+MV3.3 treatment on handwashing outcomes (Column 3 of Panel A and B). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 5—: Rational Habit Formation

	(1) Day -54 to -1	(2) Day -21 to -1	(3) Day -54 to -1	(4) Day -21 to -1
	Likelihood of using during reported dinnertime			
Anticipated triple incentive (IV 3.1)	-0.0051	-0.0381		
	(0.0235)	(0.0306)		
Anticipated monitoring (MV 3.1)			0.0484**	0.0763**
			(0.0243)	(0.0286)
Mean of pure control	0.4542	0.5386	0.2333	0.2085
	[0.4979]	[0.4986]	[0.4229]	[0.4063]
Comparison group				
N	23157	9036	16246	6252

Notes: Observations are at the household-day level. Robust standard errors in parentheses and clustered at the village level for all regressions. Standard deviation in brackets. All regressions include day and village level fixed effects. Comparison group for "Anticipated triple incentive" is the group that was surprised with the triple incentive on Day 0; comparison group for "Anticipated monitoring" is the group that was surprised with the monitoring service on Day 0. p-values adjusted for multiple hypothesis testing (MHT) using Anderson (2008). Two MHT families: the 2 coefficients of the effect of the IV3.1 treatment on handwashing outcomes (Columns 1-2); and the 2 coefficients of the effect of the IV3.1 treatment on handwashing outcomes (Columns 3-4). * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 6—: Preferred diarrhea and ARI measures at eight months—ITT estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Whether child had any loose stool in last two weeks		Total days of loose stool in last two weeks		Whether child showed any ARI symptoms in last two weeks		Total days of ARI in last two weeks	
Treated household	-0.0215**	-0.0250**	-0.0551**	-0.0783**	-0.0292**	-0.0419**	-0.1679**	-0.1988**
	(0.008)	(0.011)	(0.021)	(0.028)	(0.014)	(0.017)	(0.077)	(0.096)
Mean of pure control	0.099	0.099	0.207	0.207	0.269	0.269	1.245	1.245
	[0.299]	[0.299]	[0.802]	[0.802]	[0.444]	[0.444]	[2.458]	[2.458]
With baseline controls								
N	4936	3333	4951	3342	4951	3342	4951	3342

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Sample includes children younger than fourteen years. Data was collected seven to eight months after rollout. "Treated households" is any household that received a dispenser (the pooled sample of incentive, monitoring, and dispenser only households). "Whether child showed any ARI symptoms" equals one if the child experienced any of the following in the two weeks prior: runny nose, nasal congestion, cough (with or without sputum production), ear discharge, hoarseness of voice, sore throat, difficulty breathing or a prescription from a doctor for such. Baseline controls include: child age, child sex, baseline height, baseline weight, baseline mid-arm circumference, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). One MHT family: the 8 coefficients of the effect of the Household Treatment variable on the diarrhea and ARI outcomes (Columns 1-8). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table 7—: Child anthropometric outcomes after eight months—ITT estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Weight-for-age z-score		Height-for-age z-score		Mid-arm circ.-for-age z-score	
Treated household	0.1430*	0.1316*	0.2123*	0.2082*	0.0979*	0.0612
	(0.077)	(0.064)	(0.101)	(0.092)	(0.060)	(0.052)
Mean of pure control	-2.167	-2.167	-1.866	-1.866	-1.365	-1.365
	[1.087]	[1.087]	[1.573]	[1.573]	[0.990]	[0.990]
With baseline controls						
N	944	852	943	851	939	847

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Height-for-age z-score (HAZ), weight-for-age z-score (WAZ), and mid-arm circumference-for-age z-score (MAZ) are calculated using WHO anthropometric methodology. Sample is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Data was collected seven to eight months after rollout. "Treated household" is any household that received a dispenser (the pooled sample of incentive, monitoring, and dispenser only households). Baseline controls include: child age, child sex, baseline HAZ, baseline WAZ, baseline MAZ, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). One MHT family: the 6 coefficients of the effect of the Household Treatment variable on the anthropometric outcomes (Columns 1-6). * p≤ 0.10, ** p≤ 0.05, *** p≤ 0.01.

Table 8—: Child health outcomes—TOT estimates

	(1) Whether child had any loose stool in last two weeks	(2) Total days of loose stool in last two weeks	(3) Whether child showed any ARI symptoms in last two weeks	(4) Total days of ARI in last two weeks	(5) Weight-for-age z-score	(6) Height-for-age z-score	(7) Mid-arm circ. for age z-score
Regularly washes (self report)	-0.0488**	-0.1546**	-0.0727**	-0.3568**	0.2025	0.2719	0.0776
	(0.021)	(0.055)	(0.032)	(0.184)	(0.110)	(0.154)	(0.088)
Mean of pure control	0.099	0.207	0.269	1.245	-2.167	-1.866	-1.365
	[0.299]	[0.802]	[0.444]	[2.458]	[1.087]	[1.573]	[0.990]
N	3255	3263	3263	3263	826	825	821

Notes: Observations are at the child level. Robust standard errors are in parentheses and are clustered at the household level. Standard deviation in brackets. Outcome data was collected seven to eight months after rollout. Z-scores in columns 5-7 are calculated using WHO anthropometric methodology. Sample in columns 5-7 is limited to children 60 months and younger and excludes children with implausible z-scores as pre-specified in the WHO methodology. Sample in columns 1-4 include children younger than fourteen. Regression shows the treatment on the treated estimates where "treated" is either (1) a household who reports washing hands regularly during dinnertime, or (2) a household whose respondent has clean hands as judged by the enumerator, both of which are instrumented for by each of the three treatment groups (incentives, monitoring and dispenser). "Whether child showed any ARI symptoms" equals one if the child experienced any of the following in the two weeks prior: runny nose, nasal congestion, cough (with or without suputm production), ear discharge, hoarseness of voice, sore throat, difficulty breathing or a prescription from a doctor for such. All regressions include the following baseline controls: child age, child sex, baseline height, baseline weight, baseline mid-arm circumference, whether the child had a cold in the two weeks prior to baseline, whether the child had a cold in the two weeks prior to baseline, whether the child had a cough in the two weeks prior to baseline, whether the child had diarrhea in the two weeks prior to baseline, and the number of months the child was breastfed. p-values adjusted for multiple hypothesis testing (MHT) using Benjamini et al. (2006). Two MHT families: the 4 coefficients of the effect of the Regularly Washes variable on the diarrhea and ARI outcomes (Columns 1-4); and the 3 coefficients of the effect of the Regularly Washes variable on the anthropometric outcomes (Columns 5-7). * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

REFERENCES

Aggarwal, Shilpa, Rebecca Dizon-Ross, and Ariel D. Zucker. 2020. "Incentivizing Behavioral Change: The Role of Time Preferences." *NBER Working Papers 27079, National Bureau of Economic Research, Inc.*

Allcott, Hunt, and Judd B. Kessler. 2019. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *American Economic Journal: Applied Economics*, 11(1): 236–276.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *The American Economic Review*, 104(10): 3003–3037.

Andrews, B. R. 1903. "Habit." *The American Journal of Psychology*, 14(2): 121–149.

Auld, M. Christopher, and Paul Grootendorst. 2004. "An Empirical Analysis of Milk Addiction." *Journal of Health Economics*, 23(6): 1117–1133.

Aunger, Robert, Wolf-Peter Schmidt, Ashish Ranpura, Yolande Coombes, Peninnah Mukiri Maina, Carol Nkatha Matiko, and Valerie Curtis. 2010. "Three Kinds of Psychological Determinants for Hand-Washing Behaviour in Kenya." *Social Science & Medicine*, 70(3): 383–391.

Baltagi, Badi H., and James M. Griffin. 2002. "Rational Addiction to Alcohol: Panel Data Analysis of Liquor Consumption." *Health Economics*, 11(6): 485–491.

Banerjee, Abhijit Vinayak, Esther Duflo, Rachel Glennerster, and Dhruva Kothari. 2010. "Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives." *British Medical Journal*, 340: C2220.

Barker, J., I. B. Vipond, and S. F. Bloomfield. 2004. "Effects of Cleaning and Disinfection in Reducing the Spread of Norovirus Contamination via Environmental Surfaces." *Journal of Hospital Infection*, 58(1): 42–49.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96(4): 675–700.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *The American Economic Review*, 81(2): 237.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1994. "An Empirical Analysis of Cigarette Addiction." *The American Economic Review*, 84(3): 396–418.

Bennett, Daniel, Asjad Naqvi, and Wolf-Peter Schmidt. 2018. "Learning, Hygiene and Traditional Medicine." *The Economic Journal*, 128(612): F545–F574.

Cameron, Dr. Samuel. 2000. "Nicotine Addiction and Cigarette Consumption: A Psycho-Economic Model." *Journal of Economic Behavior & Organization*, 41(3): 211–219.

Cameron, Lisa, Manisha Shah, and Susan Olivia. 2013. *Impact Evaluation of a Large-Scale Rural Sanitation Project in Indonesia. Policy Research Working Papers*, The World Bank.

Chaloupka, Frank. 1991. "Rational Addictive Behavior and Cigarette Smoking." *Journal of Political Economy*, 99(4): 722–742.

Charness, Gary, and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica*, 77(3): 909–931.

Chase, Claire, and Quy-Toan Do. 2012. "Handwashing Behavior Change at Scale: Evidence from a Randomized Evaluation in Vietnam." The World Bank.

Clasen, Thomas, Sophie Boisson, Parimita Routray, Belen Torondel, Melissa Bell, Oliver Cumming, Jeroen Ensink, Matthew Freeman, Marion Jenkins, Mitsunori Odagiri, Subhajyoti Ray, Antara Sinha, Mrutyunjay Suar, and Wolf-Peter Schmidt. 2014. "Effectiveness of a Rural Sanitation Programme on Diarrhoea, Soil-Transmitted Helminth Infection, and Child Malnutrition in Odisha, India: A Cluster-Randomised Trial." *The Lancet Global Health*, 2(11): e645–e653.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *The American Economic Review*, 100(1): 35–69.

Duhigg, Charles. 2012. *The Power of Habit: Why We Do What We Do in Life and Business.* New York:Random House Publishing Group.

Dupas, Pascaline. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment." *Econometrica*, 82(1): 197–228.

Dupas, Pascaline, and Edward Miguel. 2016. "Impacts and Determinants of Health Levels in Low-Income Countries."

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American economic journal. Applied economics*, 8(4): 160–188.

Galiani, Sebastian, Paul Gertler, Nicolas Ajzenman, and Alexandra Orsola-Vidal. 2016. "Promoting Handwashing Behavior: The Effects of Large-Scale Community and School-Level Interventions." *Health Economics*, 25(12): 1545–1559.

Gruber, Jonathan, and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *The Quarterly Journal of Economics*, 116(4): 1261–1303.

Haggerty, Patricia A., Kalengaie Muladi, Betty R. Kirkwood, Ann Ashworth, and Manwela Manunebo. 1994. "Community-Based Hygiene Education to Reduce Diarrhoeal Disease in Rural Zaire: Impact of the Intervention on Diarrhoeal Morbidity." *International Journal of Epidemiology*, 23(5): 1050–1059.

Han, Aung Myo, and Thein Hlaing. 1989. "Prevention of Diarrhoea and Dysentery by Hand Washing." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 83(1): 128–131.

Hoddinott, John, Jere R. Behrman, John A. Maluccio, Paul Melgar, Agnes R. Quisumbing, Manuel Ramirez-Zea, Aryeh D. Stein, Kathryn M. Yount, and Reynaldo Martorell. 2013. "Adult Consequences of Growth Failure in Early Childhood." *The American Journal of Clinical Nutrition*, 98(5): 1170–1178.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2020. "Replication Data for Rational Habit Formation: Experimental Evidence from Handwashing in India." *American Economic Association [publisher and distributor]. https://doi.org/10.3886/E117401V1.*

Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2015. "The Persistence of Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand."

Iyer, P., and J. Sara. 2005. "The Handwashing Handbook: A Guide for Developing a Hygiene Promotion Program to Increase Handwashing with Soap." *The World Bank.*

James, William. 1914. *Habit.* New York:H.Holt and company.

Jessoe, Katrina, and David Rapson. 2014. "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use." *American Economic Review*, 104(4): 1417–1438.

Kremer, Michael, and Alix Peterson Zwane. 2007. "Cost-Effective Prevention of Diarrheal Diseases: A Critical Review." *SSRN Electronic Journal.*

Laibson, David. 2001. "A Cue-Theory of Consumption." *The Quarterly Journal of Economics*, 116(1): 81–119.

Luby, Stephen P, Mahbubur Rahman, Benjamin F Arnold, Leanne Unicomb, Sania Ashraf, Peter J Winch, Christine P Stewart, Farzana Begum, Faruqe Hussain, Jade Benjamin-Chung, Elli Leontsini, Abu M Naser, Sarker M Parvez, Alan E Hubbard, Audrie Lin, Fosiul A Nizame, Kaniz Jannat, Ayse Ercumen, Pavani K Ram, Kishor K Das, Jaynal Abedin, Thomas F Clasen, Kathryn G Dewey, Lia C Fernald, Clair Null, Tahmeed Ahmed, and John M Colford. 2018. "Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on Diarrhoea and Child Growth in Rural Bangladesh: A Cluster Randomised Controlled Trial." *The Lancet Global Health*, 6(3): e302–e315.

Luby, Stephen P, Mubina Agboatwalla, Daniel R Feikin, John Painter, Ward Billhimer, Arshad Altaf, and Robert M Hoekstra. 2005. "Effect of Handwashing on Child Health: A Randomised Controlled Trial." *The Lancet*, 366(9481): 225–233.

Marshall, Alfred. 1890. *Principles of Economics.* Macmillan and Company.

McKay, Sue, Estelle Gaudier, David I. Campbell, Andrew M. Prentice, and Ruud Albers. 2010. "Environmental Enteropathy: New Targets for Nutritional Interventions." *International Health*, 2(3): 172–180.

Neal, David, Jelena Vujcvic, Orlando Hernandez, and Wendy Wood. 2015. "The Science of Habit: Creating Disruptive and Sticky Behavior Change in Handwashing Behavior." *Washington, D.C.: USAID/WASHplus Project.*

Null, Clair, Christine P Stewart, Amy J Pickering, Holly N Dentz, Benjamin F Arnold, Charles D Arnold, Jade Benjamin-Chung, Thomas Clasen, Kathryn G Dewey, Lia C H Fernald, Alan E Hubbard, Patricia Kariger, Audrie Lin, Stephen P Luby, Andrew Mertens, Sammy M Njenga, Geoffrey Nyambane, Pavani K Ram, and John M Colford. 2018. "Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on Diarrhoea and Child Growth in Rural Kenya: A Cluster-Randomised Controlled Trial." *The Lancet Global Health*, 6(3): e316–e329.

O'Donoghue, Ted. 2001. "Addiction and Present-Biased Preferences." 53.

Pollak, Robert A. 1970. "Habit Formation and Dynamic Demand Functions." *Journal of Political Economy*, 78(4): 745–763.

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics*, 7(3): 51–84.

Sanderson, P. J., and S. Weissler. 1992. "Recovery of Coliforms from the Hands of Nurses and Patients: Activities Leading to Contamination." *Journal of Hospital Infection*, 21(2): 85–93.

Spinnewyn, Frans. 1981. "Rational Habit Formation." *European Economic Review*, 15(1): 91–109.

Stigler, George J., and Gary S. Becker. 1977. "De Gustibus Non Est Disputandum." *The American Economic Review*, 67(2): 76–90.

Taubinsky, Dmitry. 2014. "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice."

WHO. 2006. "WHO Child Growth Standards Based on Length/Height, Weight and Age." *Acta Paediatrica (Oslo, Norway: 1992). Supplement*, 450: 76–85.

WHO. 2009. *WHO Guidelines on Hand Hygiene in Health Care.* Geneva, Switzerland:World Health Organization, Patient Safety.

Wishnofsky, Max. 1958. "Caloric Equivalents of Gained or Lost Weight." *The American Journal of Clinical Nutrition*, 6(5): 542–546.

Wood, Wendy, and David T. Neal. 2007. "A New Look at Habits and the Habit-Goal Interface." *Psychological Review*, 114(4): 843–863.