Workfare, Wellbeing and Consumption:

Evidence from a Field Experiment with Kenya's Urban Poor

Syon P. Bhanot[*], Jiyoung Han[†], Chaning Jang[‡]

December 14, 2016

Abstract

Workfare and vouchers are often used in social welfare systems, but little empirical evidence exists on their impact on wellbeing. We conducted a randomized experiment in Kenya, and tested two elements of social welfare design: workfare versus welfare and restricted versus unrestricted vouchers. Participants were randomly assigned to a "Work" condition, involving work for unrestricted vouchers, or one of two "Wait" conditions, involving waiting for vouchers that were either unrestricted or restricted to staple foods. We find that working significantly improved psychological wellbeing relative to waiting. Furthermore, although the restrictions were infra-marginal, partially restricted vouchers crowded-in spending on staple foods.[1]

1 Introduction

It is important to understand how the structure of social welfare programs influences the wellbeing and decision-making of their target populations. A recent body of academic research in development has focused on the impact of low-cost interventions on wellbeing (Cohen et al., 2012; Banerjee et al., 2011), but there is less concrete evidence on how the design of social welfare programs influences key attitudes and behaviors. In this paper, we contribute experimental evidence in two main areas where existing work is limited. First, we provide some of the first causal evidence around the effect of "workfare" (welfare programs that require beneficiaries to work for welfare payments) on the wellbeing and decisions of target populations. In doing so, we build on behavioral economics concepts of "mental accounting," which suggest that earned money may be treated differently than unearned money (Thaler, 2008). Second, while there is a large body of research on transfers in development (Blattman and Niehaus, 2014; Lagarde et al., 2007; Gertler, 2004; Fiszbein et al., 2009), there is less experimental research on how restrictions in spending built into transfers in developing contexts (for example through vouchers targeting specific areas of spending) might influence consumption decisions (Aker, 2013; Hidrobo et al., 2014).

Given that publicly-available randomized evaluations are uncommon among institutional bodies that oversee welfare programs, we conducted a randomized field experiment that seeks to mimic real-world social welfare programs to test our hypotheses. Our intervention was conducted over a 10-day period in a low-income area in Nairobi, and tests how two specific welfare program design elements can influence decision making. Specifically, we explore: 1) the effect of workfare versus welfare, or "working for" rather than simply receiving welfare benefits as a "handout," on spending behavior and psychological wellbeing; and 2) the effect of restricted (limited for use on a specific subset of items, namely food "staples") versus unrestricted vouchers on spending behavior. We also randomize the visual framing of the vouchers themselves, but find minimal impacts from these minor manipulations (and therefore refrain from an extensive discussion of this aspect in this paper).

We conducted the experiment through the Busara Center for Behavioral Economics in Nairobi, Kenya, and the PBK Nonic Supermarket in Kawangware, an informal settlement in Nairobi. We recruited 432 participants using a random draw from the Busara subject pool in Kawangware. Participants were instructed to show up to one of three experiment locations in Kawangware for one hour every day, for ten consecutive weekdays. Of the 432 participants recruited, 383 showed up for at least one of the ten sessions, and 263 showed up for all ten sessions as instructed.

Participants were randomized into three experimental treatment arms, hereafter referred to as the "Work," "Wait-Unrestricted," and "Wait-Hybrid" treatments. Each treatment was conducted at one of the three experiment locations, and remained there for the entirety of the experiment. Each day, participants in the Work treatment separated rice and lentils from a mix in exchange for two unrestricted PBK grocery vouchers, each worth KES 100 (KES 200 in total).[2] Meanwhile, participants in the Wait-Unrestricted treatment were at their experiment location for the same period of time as the Work treatment participants, and received the same compensation of two KES 100 unrestricted PBK grocery vouchers per daily session. However, they sat idle rather than working during their time at the experiment location. Finally, participants in the Wait-Hybrid treatment received the same on-site treatment as those in the Wait-Unrestricted treatment, but received one unrestricted PBK voucher (worth KES 100) and one restricted PBK voucher (worth KES 100) per day. The restricted voucher could only be used for the purchase of "staple foods," namely maize/wheat flour, rice, sugar, and cooking oil.

We measured two primary sets of outcomes for each individual in the experiment. First, we tracked participants' spending behavior as they redeemed their grocery vouchers, using individual receipts collected at the supermarket. Second, we measured psychological wellbeing in two ways — through daily PANAS surveys (Watson et al., 1988) on affective state, administered at the beginning of each session, and an endline survey with broader wellbeing and attitude measures.

The experimental conditions were designed to provide insights on two primary elements of social welfare program design. First, comparing participant expenditures and psychological wellbeing in the Work and Wait-Unrestricted treatments provides an estimate of the effect of working for payment rather than receiving a welfare "handout." Second, comparing spending outcomes in the Wait-Unrestricted and Wait-Hybrid treatments provides an estimate of the effect of partial voucher restrictions on spending decisions.

There are two main results from the experiment. First, receiving vouchers in exchange for working had large and positive impacts on psychological wellbeing and perceptions of self-efficacy, relative to waiting idly in exchange for the same benefit. Subjects in the Work treatment in particular reported consistently higher psychological wellbeing than those in the Wait-Unrestricted treatment on nearly all measures. We did not find an effect of the work treatment on spending decisions. Second, we find that participants in the Wait-Hybrid group, who received half of their payment as restricted "staples" vouchers, spent significantly more on staple foods than participants in the Wait-Unrestricted group. This result is interesting because Wait-Unrestricted participants also spent, on average, a little over half of their voucher money on the same

---

[2]KES 100 = USD 1.13 on September 22, 2014. As of June 2015, the government-set minimum wage for an unskilled urban worker was KES 5436 a month, or approximately KES 180 a day. Therefore the voucher amounts were within range of daily wages typically found in an urban Kenyan setting like Kawangware.

staple foods, suggesting that the voucher restrictions were infra-marginal for the Wait-Hybrid group. Further analysis of this difference suggests that the decision to spend more on staple foods by those in the Wait-Hybrid group appears inconsistent with a standard model of a utility-maximizing consumer. Instead, we argue that our findings are evidence of a "flypaper effect" (Hines and Thaler, 1995) from restricted vouchers, whereby money provided to individuals for specific goods "sticks where it hits," increasing overall expenditure on these goods.

In addition to the academic implications, these results provide three important policy insights . First, if the individual and social gains from workfare payments outweigh those accrued through standard welfare schemes, it suggests that workfare might represent an efficient and welfare-enhancing way of redistributing goods and money to low-income individuals. Second, this study offers new evidence that the way in which a welfare program is administered may have important impacts on individual wellbeing and consumption choice beyond the simple income effect from receiving the benefit. Third, our lab-to-field methodology allows us to study the effects of different welfare program features on an under-examined developing country population in a quasi-real-world setting. This study therefore can inform the design of welfare programs actively under consideration by the government of Kenya, and by other similar nations in the developing world.

The paper proceeds as follows. Section 2 provides background, outlines existing literature, and provides policy context for this experiment. Section 3 outlines the experiment itself. Section 4 offers a brief outline of the data collected during the experiment. We then discuss the empirical strategy in Section 5, and present the results in Section 6. Section 7 provides a discussion, with recommendations for further research.

2 Background

2.1 Existing Literature on Workfare and Welfare, and Related Concepts from Behavioral Economics

The relationship between work and welfare has received attention in both the policy and academic spheres. Our experiment explores questions in this area, drawing on work in psychology, public economics, and behavioral economics on motivation, social program targeting, and mental accounting, as well as their implications for wellbeing and consumer theory.

There are a number of open questions regarding how recipients of social welfare are impacted by such programs. Our focus on "workfare"-style programs relates to academic literature on the role of "ordeals" in the provision of social welfare. Specifically, (Nichols and Zeckhauser, 1982) argue that subjecting welfare

4

recipients to ordeals enables more effective targeting of benefits to the most needy, who are most willing to endure the opportunity cost imposed by the ordeal in exchange for the benefit. Ordeals often come in the form of a non-monetary price (like the opportunity cost imposed by an extended wait in line), which can be an efficient approach when price adjustments are ineffective (Cohen et al., 2012).

Ordeals often take the form of work, which has the added theoretical benefit of increasing economic output and decreasing government expenditure (Blumkin et al., 2010). In practice, large public works programs like India's National Rural Employment Guarantee Act (NREGA) or Peru's Trabajo programs do not view work as an ordeal, but rather as a right and means of social protection – these programs offer wage employment to qualifying individuals, typically with an aim to improve livelihoods and wellbeing. A growing body of literature supports this view, arguing that an individual's intrinsic motivation and a sense of higher purpose can encourage greater productivity and happiness in existing jobs (Schwartz, 2015; Pfeffer, 1998). This implies that "workfare" itself may generate utility for the individual beyond material remuneration. While jobs can thus be a source of utility, other studies show that people may value being occupied by work as a means to avoid the disutility of being idle (Hsee et al., 2010). However, to our knowledge, there is no literature that directly compares the impact of workfare and idle waiting on welfare beneficiaries' psychological wellbeing. We offer evidence on this question. Furthermore, our comparison of the Work and Wait-Unrestricted treatment groups allows us to explore whether the nature of the ordeal itself — not the equivalent time cost it imposes — influences behavioral response.

The topic of how behavior differs under workfare versus welfare programs also touches on mental accounting, a key aspect of behavioral economics that has received sustained attention in the last few decades. Mental accounting refers to the phenomenon whereby individuals code, categorize, and assess spending in ways that violate the axioms of rational economic decision making (Thaler, 1999). For example, mental accounting models posit that people tend to group money into categories — such as groceries, rent, or luxuries — and spend money according to these implicit budgets. This compartmentalization of money may also drive people to treat money differently based on how it was acquired, which is not consistent with a rational model. There is existing evidence of the empirical validity of this concept in real-world settings. For example, Arkes et al. (1994) find that people have a greater marginal propensity to consume from "windfall" earnings than money earned through work, while Beshears and Milkman (2009) find that people are more likely to spend windfall money on non-routine purchases. In the developing country context, one experimental study found that rural Tanzanians are more likely to spend earned money on basic consumption goods or education, and unearned money on non-basic consumption goods such as alcohol, tobacco, or non-staple foods (Pan and Christiaensen, 2012). Additionally, Davies et al. (2009) find that low-income Malawian households are more

5

likely to allocate remittances to education, suggesting that households make spending decisions based on the source of funds.

In our experiment, we explore whether individuals treat money differently based on how it was obtained. Specifically, we compare spending behavior in the Wait-Unrestricted and Work experimental conditions to determine if "earning" versus "getting" money results in different consumption decisions.

2.2 Existing Work on Restricted Transfers

Traditional microeconomic theory posits that individuals make consumption choices to maximize their utility, subject to their budget constraints. Our experiment tests one aspect of the standard model challenged by mental accounting, namely that rational individuals treat money as fungible. In particular, this experiment provides an individual-level test of the "flypaper effect," which suggests that money that has been ex-ante labeled, or directed at a specific category of spending, may "stick where it hits" (Hines and Thaler, 1995), increasing expenditure on items associated with that mental account beyond what it might in a rational model. For example, a rational individual receiving a $50 gift card to a store that she frequents regularly would "substitute" $50 of planned expenditure at that store with that gift card, in effect "freeing up" $50 in cash to be spent across all categories of spending. However, a mental accounter subject to the flypaper effect would respond to the $50 gift card by increasing expenditure at the store by closer to $50 (or at least by more than the rational individual would). There is some evidence of this effect in real world decision making contexts, including in financial portfolio choice among U.S. households (Choi et al., 2007), but evidence of its broad applicability is limited.

It is especially important to study the flypaper effect in the context of social welfare programs. Governments and other benefits providers often seek to achieve their own ends by directing consumers towards particular choices with in-kind transfers. For example, the U.S. Supplemental Nutrition Assistance Program (SNAP) restricts benefits to particular types of goods (Currie and Gahvari, 2008). Research by Hoynes et al. (2014) on the effect of SNAP on consumption choices finds that SNAP benefits are infra-marginal for the majority of eligible households, and that those households should treat the benefits like a cash-transfer. Furthermore, there is an ongoing debate on the usefulness of in-kind transfers relative to cash transfers in the developing country context (Devereux, 2006), and recent research has begun to address this question. For example, Hidrobo et al. (2014) present an experiment comparing the effect of in-kind transfers and cash in Ecuador. While all treatments in their study (cash, food vouchers, and food transfers) improved the quality and quantity of food consumed, vouchers restricted to food led to the most dietary diversity and were the most cost-effective, while pure in-kind food transfers led to the greatest increase in caloric intake.

Our experiment provides some of the first experimental evidence on the flypaper effect in individual consumption choice. Specifically, our design builds on that in Hidrobo et al. (2014) by providing subjects with a combination of restricted and unrestricted "near-cash" grocery vouchers that were not entirely fixed to food items. This setup allows for a direct test of the flypaper effect (by exploring how spending on items in the restricted vouchers differed in the treatment groups). Additionally, our design seeks to emulate a scalable in-kind welfare scheme that provides recipients with some restriction in allowed expenditure but still ample choice. Although consumption choices in our experiment were limited to the goods available at a specific grocery store, past surveys indicate that individuals in our subject pool's income group tend to allocate a majority of their expenditures to basic consumption goods that could be found in such a store (Amendah et al., 2014).

3 Experiment Overview

3.1 Experiment Design

Timing, Participants, and Study Sites

The experiment took place over ten consecutive business days during the two week period from September 22, 2014 to October 3, 2014. The participants in this experiment were 432 individuals living in the Kawangware area of Nairobi, Kenya, selected at random for recruitment to the study from the subject pool maintained by the Busara Center for Behavioral Economics.[3] Participants were initially contacted by phone and asked if they were available and willing to participate in a ten-day study. Participants were then randomly assigned to three treatment groups, described in detail below, located in three different community halls within the Kawangware settlement. Each community hall hosted four sessions per day at the same times each day: 9:30 AM, 11:30 AM, 1:30 PM, and 3:30 PM.[4] Each hall was 5-10 minutes, by foot, from the PBK Nonic Supermarket, but far enough from one another to limit the potential for contamination between treatment groups. A map depicting the location of the treatment sessions and the PBK Nonic Supermarket can be found in the online appendix.

---

[3]Busara recruits study participants periodically throughout the year from different areas of Nairobi, including university students, residents of Kibera, and residents of Kawangware. Busara and local community liaison officers engaged in a large recruitment effort in September 2014 to recruit participants for this study. Of those registered into the database, Busara randomly selected a subset to invite to the study.

[4]After participants were recruited by phone, they were asked to select one of four time slots during the day for their participation in the study. Once participants were confirmed for a given time slot, they were randomly assigned to one of the three treatment groups and notified of their experiment location. They were not notified of the specifics of the experiment, or that there were more than one session, location, or treatment either before, during or after the study. This design mitigated possible selection bias stemming from the time of day that people were available to participate or from selection on location.

Treatment Groups

Participants were randomized into three treatment groups, namely:

1. Wait-Unrestricted: Participants were asked to wait in the treatment location for one hour each day in exchange for a payment of two unrestricted PBK Nonic Supermarket vouchers worth KES 100 each (KES 200 in total). This serves as our de-facto comparison group.

2. Work: Participants were asked to complete a work task in the treatment location for one hour each day in exchange for a payment of two unrestricted PBK Nonic Supermarket vouchers worth KES 100 each (KES 200 in total).

3. Wait-Hybrid: Participants were asked to wait in the treatment location for one hour each day in exchange for a daily payment of one unrestricted PBK Nonic Supermarket voucher worth KES 100, and one restricted PBK Nonic Supermarket voucher worth KES 100 that could be used for staple food items only (Maize/wheat flour, rice, sugar, or cooking oil). We refer to this as the "staples" voucher.

The face value of the vouchers provided each day to the participants was constant across treatments (KES 200), and was an amount slightly higher than an average daily wage for most participants. The experimental procedure for each treatment is outlined below.

Treatment 1: Wait-Unrestricted

Individuals were asked to arrive at the Urumwe Youth Group hall to collect vouchers every day for ten weekdays. Upon arriving at the hall, field officers asked participants to wait for their vouchers without providing any concrete justification for the delay. Field officers allowed participants to bring reading materials or other diversions if they wished. Occasionally, field officers would ask participants to come to the front of the room to confirm their name, or write their name/signature on a sheet of paper. This was done to mimic the tedium of a traditional welfare "ordeal" process, as outlined in Nichols and Zeckhauser (1982). This also equalized the time that participants spend in the experiment across treatments. Field officers were present at all times during the waiting period. After an hour of waiting, participants were called to the front by name and given two unrestricted PBK Nonic Supermarket vouchers worth KES 100 each.

Treatment 2: Work

Individuals were asked to arrive at the Kabiro Social Hall for one hour every day for ten weekdays. After checking in with field officers, they were required to separate rice and lentils from a mix into small cups. Field officers distributed large plastic cups filled with the rice and lentil mix to all participants at the beginning

of every session and two smaller plastic cups for participants to collect the separated grains. They were not told precisely why they were doing this work task, and anecdotally few asked field officers for an explanation of any kind.

All participants in the sessions had space to work. Past studies with this population suggested that men might find doing such work in the presence of females embarrassing, so to avoid any distortionary effects, the men and women in the sample were separated in the hall. Field officers recorded the time when each participant started and stopped sorting grains. At the end of the work period, field officers would note the session conclusion time and measure the weight of rice and lentils that were sorted by each participant. Field officers were present at all times during the work sessions and participants knew that their output would be weighed. At the end of each day's work, participants were given two unrestricted PBK Nonic Supermarket vouchers worth KES 100 each (regardless of productivity).

<u>Treatment 3: Wait - Hybrid</u>

The Wait-Hybrid treatment, which took place at the Kawangware Day Nursery School, was identical to the Wait-Unrestricted treatment in every way, except that instead of receiving two unrestricted PBK Nonic Supermarket vouchers, participants received one KES 100 unrestricted voucher and one KES 100 restricted voucher, which could be used for staples foods only.

<u>Cross randomization of framing effects</u>

To test framing effects, participants were randomly assigned different messaging on their vouchers. There were two types of voucher messaging: the "Self" voucher, which had an image of a woman with money and a message in text beneath the image that read, "Treat yourself to something nice!"; and the "Family" voucher, which had an image of a family eating together and a message in text beneath the image that read, "Be responsible! Bring home something for the family!" Example vouchers can be found in the online appendix.

Participants were randomized into two groups within treatments — half of each treatment group received the Self voucher for the first five days of the study and the Family voucher for the last five days, and vice-versa for the other half of participants. The messaging on the vouchers was briefly reinforced by field officers, particularly for illiterate participants. However, the field officers clarified that participants were not required to purchase any items or types of items in particular.

Shopping at PBK

After receiving their vouchers, participants could spend them to purchase goods at the PBK Nonic super-market, a short walk from the study site. The supermarket sells a wide variety of food, household supplies,

school supplies, and other sundries. Participants could spend their vouchers on any product at the supermarket (except for staples vouchers, which were restricted to the aforementioned goods) during normal business hours. Participants could redeem their vouchers anytime from the start of the experiment until a week after the experiment concluded. This timeline gave participants up to three full weeks to redeem their vouchers.

4 Data

4.1 Outcomes

We collect data on and explore three broad sets of outcomes, outlined below.

Outcomes I: Expenditures

During the course of the experiment, PBK accepted the vouchers distributed during the experiment and redeemed the value of the vouchers for any item (or the set of restricted items), according to the instructions on each voucher. Vouchers were labeled with ID numbers reflecting the individual's unique identifier, the treatment group, and the date of issuance in a manner that was not transparent to participants.

Every time a participant paid for his/her purchases with vouchers, PBK Nonic Supermarket staff stapled the receipt to the voucher. Next, a Busara research assistant reviewed the voucher(s) and receipt for errors. The research assistant then collected all vouchers and receipts at the end of the day and returned them to the Busara office, where they were double-entered into a database.

Through this partnership with PBK Nonic Supermarket, we were able to extract detailed data on spending by category, date, and voucher type (Self or Family, restricted or unrestricted). This data served as basis for the analysis on consumption.

Outcomes II: Baseline, Daily, and Endline Surveys

Three survey types were administered during the study. First, on the initial day of the study, participants completed a baseline survey. The baseline survey asked questions related to people's current emotions using an abbreviated 6-item PANAS scale (Watson et al., 1988), weekly spending habits, employment, household characteristics, familiarity with the PBK Nonic Supermarket, and decision-making power within the household.

Second, every day of the study after the initial day, participants completed a daily survey that contained the abbreviated PANAS questions on current emotions. These were completed at the start of the sessions.

Table 1: Days of Attendance

	10	9	8	7	6	5	4	3	2	1	0
Number	263	36	8	3	4	10	5	4	14	36	49
Percent	60.9%	8.3%	1.9%	0.7%	0.9%	2.3%	1.2%	0.9%	3.2%	8.3%	11.3%
Cumulative	60.9%	69.2%	71.1%	71.8%	72.7%	75%	76.2%	77.1%	80.3%	88.7%	100.0%

Third, at the end of the study participants completed an endline survey, which asked a series of questions on self-esteem, general happiness, and optimism. The endline survey was designed to measure overall life satisfaction and wellbeing, rather than incidental happiness, enabling us to distinguish between the effects of the treatments on affect as opposed to longer run effects on disposition. The endline survey also asked about family dynamics, income levels, how participants approached spending the vouchers, and how they felt about their consumption decisions.

The online appendix provides copies of the exact surveys administered.

Outcomes III: Attendance

Finally, for all treatment groups, attendance and timeliness to sessions was tracked by on-site field officers.

4.2   Attrition and Balance

Of the 432 participants recruited, 383 (88.7%) attended at least one of the ten sessions, and 263 (60.9%) had perfect attendance. Table 1 outlines the number of individuals who attended for each of the possible number of days.

Of the 432 recruited participants, 360 participated in the baseline survey, while 347 participated in the endline. 25 participants were surveyed at baseline but not endline, while 12 were surveyed at endline but not baseline. The 25 "attriters" were significantly more likely to be male, younger, and better educated than those that completed both baseline and endline.[5] In the next section we explore if baseline characteristics were balanced among those surveyed at baseline and endline.

Table 2 presents the results of a randomization check on baseline demographics characteristics across those participating in baseline and endline, suggestive of successful randomization on observables.

---

[5]The table assessing the differences between attriters and endline participants can be found in the online appendix, along with the full range of demographic comparisons. Another way to think about attrition in this context are "no-shows," who were invited to but never attended any experimental sessions. There were 49 "no-shows," who differ from compliers in similar ways to the "attriters;" they were generally younger, male, and with fewer children. The table on "no-shows" can also be found in the online appendix.

There is generally very little evidence of imbalance across treatments. One exception is that there were more self-reported "purchase decision-makers" in the Work treatment than in the others, both at baseline and endline. Our analysis controls for the decision-maker variable so as to correct for any potential bias.

5 Empirical Approach

In this section we outline the basic econometric approach to measuring the effect of the treatments on expenditures and wellbeing. This analysis was pre-specified and registered at the AEA Social Science Registry prior to any analysis (Bhanot et al., 2015).

5.1 Basic Specification

Our basic treatment effects specification for the primary effects of interest estimates the following equation:

$$y_i = \beta_0 + \beta_1 WORK_i + \beta_2 HYBRID_i + \delta y_{i,t=1} + \varepsilon_i \tag{1}$$

where y_i is the outcome of interest for individual i. $WORK_i$ and $HYBRID_i$ are dummy variables equal to 1 if the participant was randomly assigned to the Work or Wait-Hybrid condition, respectively, and 0 otherwise. Note that Wait-Unrestricted is the omitted group in this specification. ε_i is the unobserved error component, which is assumed to be serially uncorrelated. Where possible, we control for baseline levels of the outcome variables, $y_{i,t=1}$ to improve statistical power (McKenzie, 2012).[6]

To analyze the effect of voucher messaging (Self vs. Family) on spending and subjective well-being, we first use the following specification:

$$y_{i,t<j} = \beta_0 + \beta_1 SELF_i + \delta y_{i,t=1} + \varepsilon_i \tag{2}$$

Where $SELF_i$ is a dummy variable equal to 1 if the participant was randomly assigned to the "Self" vouchers first, and 0 if the participant was assigned to the "Family" vouchers first. $t < j$ refers to the outcomes measured in the first j days of surveys, or expenditures captured using vouchers from the first j days. Restricting $j = 5$ allows for between-group comparison to identify the effect on messaging, and controls for the fact that there may be differential response to within-group order effects.

---

[6]For individuals who are missing baseline data, we follow the dummy variable adjustment procedure from Cohen et al. (2013).

Table 2: Baseline and Endline Balance

	Surveyed at Baseline				Surveyed at Endline			
	(1) Control Mean (SD)	(2) Work	(3) Wait Hybrid	(4) N	(5) Control Mean (SD)	(6) Work	(7) Wait Hybrid	(8) N
Attrit	0.07 (0.26)	0.01 (0.03)	-0.02 (0.03)	360	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	347
Male	0.33 (0.47)	-0.10 (0.06)	-0.09 (0.06)	360	0.32 (0.47)	-0.10 (0.06)	-0.11 (0.06)	347
Age	36.02 (13.61)	0.12 (1.62)	0.49 (1.73)	360	35.70 (13.48)	1.24 (1.63)	0.93 (1.76)	347
Form 4 (education)	0.39 (0.49)	0.04 (0.06)	0.00 (0.06)	360	0.37 (0.48)	0.05 (0.06)	-0.02 (0.06)	347
Married	0.48 (0.50)	-0.02 (0.06)	-0.01 (0.07)	360	0.50 (0.50)	-0.02 (0.06)	-0.02 (0.07)	347
HH Size	4.56 (2.08)	-0.08 (0.25)	-0.03 (0.26)	360	4.52 (2.05)	0.03 (0.26)	0.10 (0.27)	335
Salaried Employee	0.18 (0.38)	0.01 (0.05)	-0.07 (0.05)	353	0.15 (0.36)	0.03 (0.05)	-0.03 (0.05)	328
Purchase Decision-maker	0.57 (0.50)	0.18 (0.06)	-0.01 (0.06)	360	0.53 (0.50)	0.21 (0.06)	-0.01 (0.07)	347
Total Expense (weekly)	4604.50 (4139.99)	-521.92 (458.85)	947.05 (496.46)	360	4454.81 (4263.85)	-735.58 (455.48)	886.31 (518.16)	347
Food Expense (weekly)	1741.52 (1180.22)	115.86 (145.30)	-7.57 (153.30)	359	1705.95 (1142.63)	138.48 (150.99)	4.39 (154.51)	334
PBK Patron	0.75 (0.43)	-0.00 (0.06)	-0.05 (0.06)	358	0.77 (0.42)	-0.03 (0.06)	-0.07 (0.06)	333
Joint test (p-value)		0.97	0.43			0.59	0.26	

Notes: Comparion of demographic characteristics for the three treatment groups. Column 1 and 5 are the mean (and standard deviation) of the control group, which is the Wait-Unrestricted group. Columns 2 and 6 are the treatment effects from the OLS regressions comparing the Work group to the Wait-Unrestricted group, while columns 3 and 7 are the treatment effects from the OLS regressions comparing the Wait-Hybrid group to the Wait-Unrestricted group. Figures in parentheses for columns 2 through 7 represent the standard error from the regression. Columns 4 and 8 are the sample sizes. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables.

We further test whether the impact of the intervention varies heterogeneously with pre-determined individual characteristics, measured at baseline and denoted by $\mathbf{X}_{i,t=0}$. The estimating equation for the differential effect of treatment for a particular characteristic uses interaction effects and is given by:

$$y_i = \beta_0 + \beta_1 WORK_i + \beta_2 HYBRID_i + \beta_4 X_{i,t=0} + \beta_5 T_i \times X_{i,t=0} + \delta y_{i,t=1} + \varepsilon_i. \tag{3}$$

where T_i is an indicator that takes a value of 1 for the individuals' treatment (Work, or Wait-Hybrid). β_5 captures the additional effect that treatment has on the measured outcome for individuals with characteristic X, above and beyond the for those without characteristic X. These results are presented in the online appendix.

5.2 Temporal Dynamics of the Treatment Effect on Affect

With daily data on psychological affect, we are able to observe how the interventions impact psychological well-being during the two-week period. In the daily data, we have outcome measures y_{it} for individual i for $t = 1, \ldots, 10$, where $t = 1$ is the measure after the first day of the intervention, and $t = 10$ is the measure on the last day. We therefore estimate the following specification, allowing for a specification of the treatment effect for each of the nine days following baseline:

$$y_{i,t=k} = \beta_0 + \sum_{k=2}^{10} \beta_1^k (WORK_i \times [t=k]) + \sum_{k=2}^{10} \beta_1^k (HYBRID_i \times [t=k]) + \delta y_{i,t=1} + \varepsilon_{it} \tag{4}$$

where $[t = k]$ is a dummy indicator for the kth day of the intervention. As before, $y_{i,t=1}$ is the measure of the outcome variable at baseline, and is included as a control to improve precision.

6 Results

6.1 Working Versus Waiting: Wellbeing

Table 3 reports the treatment effects on affective state for participants in the Work treatment, relative to the Wait-Unrestricted treatment, with controls for affect on Day 1. Effect sizes are reported in standard deviations. We see that participants in the Work treatment were significantly more likely to report being excited, proud, and alert on a near daily basis than their counterparts in the Wait-Unrestricted treatment. In general, they were also no more likely to report being upset or ashamed on a daily basis than the Wait-Unrestricted treatment. The last column presents overall effects, with all days pooled, dummy variables for

experiment day, and standard errors clustered at the individual level. These results validate the conclusions from the daily coefficients. The overall effects on excitement (+0.32 SD, $p < 0.01$), pride (+0.23 SD, $p < 0.01$), and alertness (+0.19 SD, $p < 0.01$) in particular are both statistically and practically significant. Taken together, it appears that working for unrestricted vouchers had a relatively greater positive effect on self-reported psychological wellbeing than waiting for the same vouchers. This finding implies that the material benefit of a social welfare program may not be the only source of utility for an individual beneficiary, but that the manner in which the benefit is obtained may also influence wellbeing.

Table 4 provides similar results, but for the Wait-Hybrid treatment relative to the Wait-Unrestricted treatment. This can help us assess the significance of the findings in Table 3, since we would not expect ex-ante that the differences in the Wait treatments on daily affect would be as large as those between working and not working. These results are not suggestive of a clear link between affect and whether or not vouchers were partly restricted to staple items, with only the daily excitement reports showing any notable difference. However, the overall effect of the Wait-Hybrid treatment on excitement (+0.15 SD, $p < 0.10$) is small relative to the equivalent point estimate for the Work treatment. Overall, we conclude that there is only weak support for the hypothesis that receiving partly restricted vouchers influenced wellbeing differently than receiving unrestricted vouchers.

Table 5 presents the results of an OLS regressions estimating the impact of the treatment arms on various psychological wellbeing measures at endline. Note that the Wait-Unrestricted treatment acts as the basis of comparison for the Work and Wait-Hybrid treatments. Our preferred specification includes controls for a number of demographic characteristics: age, number of children, gender, education, martial status, and if the participant self-identified as the decision maker in the family. Furthermore, we add controls for the time slot that the participant was invited to attend. When comparing across our primary treatment arms, we also control for which voucher messaging (Self or Family) a participant received first.

Our primary question was whether working for benefits had a different psychological impact than simply collecting benefits as welfare, without a work requirement. We find that participants in the Work treatment report statistically significantly higher levels of psychological wellbeing at endline than participants in the Wait-Unrestricted treatment on nearly all measures, as seen in Table 5. To test for overall significance, we create an index of the component measures based on (Anderson, 2008), reported in Table 5 as "Wellbeing Index." We also run seemingly unrelated regressions across all individual measures and test the joint significance of the Work treatment. Both are highly significant ($p < 0.01$). The significance of the effect holds both with and without controls.

Table 3: Work Treatment Panel Regressions - Daily Affect

| | Day of Treatment | | | | | | | | | |
	2	3	4	5	6	7	8	9	10	Overall Effect
Excited Today	0.46	0.61	0.34	0.39	0.30	0.31	0.38	0.46	0.02	0.32
	(0.13)	(0.12)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)	(0.15)	(0.08)
Upset Today	-0.01	-0.07	-0.38	-0.12	-0.08	-0.08	0.05	-0.19	-0.20	-0.11
	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)	(0.08)
Proud Today	0.38	0.16	0.33	0.43	0.33	0.25	0.21	0.19	0.07	0.23
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)	(0.12)	(0.12)	(0.13)	(0.08)
Alert Today	0.25	-0.03	0.23	0.10	0.20	0.43	0.35	0.37	0.03	0.19
	(0.14)	(0.13)	(0.13)	(0.15)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.08)
Ashamed Today	-0.01	0.28	-0.14	0.00	0.06	-0.03	0.19	0.06	0.03	0.05
	(0.15)	(0.14)	(0.14)	(0.15)	(0.14)	(0.15)	(0.15)	(0.14)	(0.13)	(0.08)
Happy Today	0.16	0.10	0.08	0.14	-0.08	0.17	0.10	0.08	-0.25	0.05
	(0.12)	(0.12)	(0.12)	(0.14)	(0.11)	(0.13)	(0.13)	(0.13)	(0.13)	(0.07)
Joint test (p-value)	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.03

Notes: Results from OLS regressions comparing participants in the Work condition to the Wait-Unrestricted condition on a number of affective state variables, with baseline controls. The effect sizes are reported in standard deviations. The columns show the effect of the Work treatment for a given day, relative to Wait-Unrestricted, controlling for baseline measure of outcome variable at the individual level. "Overall Effect" is the result of a separate regression pooling across all days, with day dummies and clustered standard errors on the individual level. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables for a given day.

Table 4: Wait-Hybrid Treatment Panel Regressions - Daily Affect

					Day of Treatment					
	2	3	4	5	6	7	8	9	10	Overall Effect
Excited Today	0.49	0.47	0.12	0.18	0.05	-0.07	0.07	0.25	0.03	0.15
	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.15)	(0.09)
Upset Today	0.04	0.02	-0.17	-0.07	0.03	-0.02	0.22	-0.01	-0.12	-0.01
	(0.13)	(0.14)	(0.13)	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)	(0.08)
Proud Today	0.15	-0.09	0.19	0.19	0.25	-0.02	0.14	-0.08	-0.06	0.07
	(0.13)	(0.14)	(0.13)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.13)	(0.08)
Alert Today	0.11	-0.14	0.15	0.01	0.11	0.21	0.04	0.17	0.02	0.06
	(0.14)	(0.14)	(0.14)	(0.15)	(0.13)	(0.13)	(0.13)	(0.14)	(0.13)	(0.08)
Ashamed Today	-0.01	-0.19	-0.03	-0.11	0.16	0.15	0.12	-0.11	0.14	0.02
	(0.15)	(0.15)	(0.14)	(0.16)	(0.14)	(0.15)	(0.15)	(0.15)	(0.13)	(0.09)
Happy Today	0.24	0.11	0.09	0.20	-0.03	0.04	-0.01	-0.02	-0.19	0.04
	(0.12)	(0.13)	(0.13)	(0.14)	(0.11)	(0.13)	(0.13)	(0.14)	(0.13)	(0.07)
Joint test (p-value)	0.05	0.02	0.68	0.78	0.23	0.30	0.35	0.74	0.33	0.09

Notes: Results from OLS regressions comparing participants in the Wait-Hybrid condition to the Wait-Unrestricted condition on a number of affective state variables, with baseline controls. The effect sizes are reported in standard deviations. The columns show the effect of the Wait-Hybrid treatment for a given day, relative to Wait-Unrestricted, controlling for baseline measure of outcome variable at the individual level. "Overall Effect" is the result of a separate regression pooling across all days, with day dummies and clustered standard errors on the individual level. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables for a given day.

Table 5: Basic Specification - Psychological Wellbeing at Endline

	(1) Control Mean (SD)	(2) Work	(3) Work	(4) Wait Hybrid	(5) Wait Hybrid	(6) N
Makes Purchase Decisions	0.70	0.11	0.02	−0.03	−0.03	347
	(0.46)	(0.06)	(0.05)	(0.06)	(0.05)	
Say in HH Decisions Increased	0.33	0.44	0.43	0.39	0.40	347
	(0.47)	(0.06)	(0.06)	(0.06)	(0.07)	
Happy with Purchases	0.00	0.33	0.29	0.21	0.15	342
	(1.00)	(0.11)	(0.09)	(0.12)	(0.13)	
Purchase Item for Self	0.43	0.23	0.18	0.34	0.30	343
	(0.50)	(0.06)	(0.07)	(0.06)	(0.06)	
Improved Self-esteem	0.80	0.13	0.14	0.07	0.08	347
	(0.40)	(0.04)	(0.05)	(0.05)	(0.05)	
Own Happiness	−0.00	0.17	0.19	−0.01	0.00	345
	(1.00)	(0.11)	(0.11)	(0.13)	(0.13)	
Comparative Happiness	0.00	0.36	0.41	0.20	0.22	345
	(1.00)	(0.12)	(0.13)	(0.14)	(0.14)	
Consistently Happy	0.00	0.39	0.34	0.45	0.45	346
	(1.00)	(0.11)	(0.11)	(0.12)	(0.12)	
Consistently Unhappy	0.00	0.25	0.30	0.67	0.67	347
	(1.00)	(0.13)	(0.13)	(0.12)	(0.13)	
Mental Health	0.00	0.31	0.29	0.07	0.08	346
	(1.00)	(0.12)	(0.11)	(0.14)	(0.14)	
Life Prospects	−0.00	0.38	0.40	0.38	0.43	343
	(1.00)	(0.11)	(0.12)	(0.12)	(0.12)	
Life Security (Women only)	0.00	0.42	0.47	0.63	0.75	258
	(1.00)	(0.15)	(0.15)	(0.16)	(0.16)	
Wellbeing Index (Weighted Avg)	−0.00	0.67	0.60	0.29	0.32	347
	(1.00)	(0.11)	(0.11)	(0.13)	(0.13)	
Joint test (p-value)		0.00	0.00	0.00	0.00	
Includes controls		No	Yes	No	Yes	

Notes: Results from OLS regressions comparing participants by treatment on endline psychological wellbeing measures. Column 1 is the mean (and standard deviation) of the control group, which is the Wait-Unrestricted group. Outcomes with mean 0.00 and standard deviation of 1 are standardized effect sizes. The other outcomes are binary. The columns show the effect of the Work (columns 2 and 3) and Wait-Hybrid (columns 4 and 5) treatments, relative to Wait-Unrestricted, with and without controls. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables.

Notably, relative to the Wait-Unrestricted group, participants in the Work treatment more strongly agreed that their say in household decisions increased (+0.43 SD, $p < 0.01$), that they were happy with their purchases (+0.29 SD, $p < 0.01$), and that they had improved self-esteem (+0.14 SD, $p < 0.01$). Those in the Work treatment also show large increases in comparative happiness, mental health, life prospects, and life security (women only), relative to the Wait-Unrestricted group ($p < 0.01$ for all). Curiously, we do not see a strong effect on self reported own-happiness, despite the fact that comparative happiness is 0.41 SD higher for those in the Work treatment than those in the Wait-Unrestricted treatment. However, this matches the similar null-effect on incidental daily happiness from the daily analysis.

However, we should note that participants in the Wait-Hybrid treatment also exhibited higher self-reported scores than those in the Wait-Unrestricted treatment on a number of endline wellbeing measures. Specifically, Table 5 shows that participants in the Wait-Hybrid treatment reported greater life security (women only, +0.75 SD, $p < 0.01$), life prospects (+0.43 SD, $p < 0.01$), and say in household decision making (+0.40 SD, $p < 0.01$) at endline than those in the Wait-Unrestricted treatment.

In light of this, how do we interpret these endline results? First, the significance of the coefficients for both the Wait-Hybrid and Work conditions relative to the Wait-Unrestricted, for those measures where they appear jointly (life security, life prospects, and household decision making), suggest caution is required when claiming that working was the reason for the boosts in those measures. These effects in endline, one might argue, could be the result of abnormal results in the Wait-Unrestricted group. That said, the fact that both the Wait-Hybrid and the Work treatments have significant coefficients is not evidence that work did not have an impact on endline wellbeing, as the Wait-Hybrid condition may have influenced these measures through a different channel (for example, receiving restricted vouchers may have empowered women in the household by binding the hands of the household in terms of decisions, improving life security, life prospects, and perceptions of household decision making). Second, we note that the magnitude of the endline and daily affect effects for the Work treatment are generally larger than for the Wait-Hybrid treatment, suggesting that the effects we observe are indicative of a genuine association between wellbeing and work.

6.2 Working Versus Waiting: Consumption

Table 6 shows the results of OLS regressions that compare the relationship between the treatment arms and both specific and broad categories of grocery purchases. We begin by comparing the Work and Wait-Unrestricted treatments (the latter is the omitted comparison group in Table 6). The results do not suggest that there were very large differences in specific categories of consumption spending between the Work and Wait-Unrestricted conditions. The point estimates do provide some weak evidence that participants in the

Work condition purchased more food (KES 104.21 KES, not significant at the 10% level) and less non-food (KES 124.54, significant at the 5% level) relative to the Wait-Unrestricted condition. However, the joint-test of coefficients for the Work condition on consumption categories, relative to the Wait-Unrestricted condition, results in a p-value of 0.08, meaning we cannot confidently assert that consumption differed amongst individuals in the two conditions. Importantly, while we cannot rule out a relationship, we do not find strong support for the findings in other developing country contexts that suggest that earned money is more readily spent on basic consumption goods (Pan and Christiaensen, 2012).

6.3 Partial Voucher Restriction: Consumption

To test the hypotheses around mental accounting and the flypaper effect in social welfare vouchers, we can assess differences in consumption between the Wait-Hybrid and Wait-Unrestricted groups in Table 6. The results suggest that participants in the Wait-Hybrid treatment – the only participants with restricted vouchers – spent KES 282.22 more on staple foods than participants in the Wait-Unrestricted treatment (who spent KES 1061.49 on staples). This result is strongly significant ($p < 0.01$). Notably, the increased purchasing of staple items by those in the Wait-Hybrid condition seems to have crowded out the purchase of non-food items (KES 251.74, $p < 0.01$). So it is not the case that the greater staples food spending in the Wait-Hybrid condition represents a reallocation of the budget from one food category (non-staple foods) to another (staple foods). This is a significant result in the low-income setting, since the provision of the staples voucher lead to greater food spending overall. This suggests targeted vouchers may be a useful a policy tool to increase spending on specific foods, without otherwise affecting intended food purchases.

This increase in staple food spending is not immediately surprising, as the greater staples spending in the Wait-Hybrid treatment could be interpreted as an entirely rational "move to the kink point," from the traditional consumer theory model of vouchers. However, this rational interpretation of the result is called into question by the high level of staple spending in the Wait-Unrestricted treatment. As column 1 in Table 6 shows, participants who received only unrestricted vouchers still spent over half of their voucher money on staple items. So these staple items were ones that participants overall want to buy in high quantities in the absence of voucher restrictions. This is suggestive evidence of a flypaper effect, whereby the restricted staples voucher might have encouraged increased staples spending overall rather than a rational reallocation of budget across consumer items.

However, it is possible that the effect we observe is partly the result of some *subset* of Wait-Hybrid participants "moving to the kink point" in staples consumption (namely, those who would not have spent much on staples in the absence of any voucher restrictions), while others in the Wait-Hybrid treatment (high staples

Table 6: Basic Specification - Consumption Totals

	(1) Control Mean (SD)	(2) Work	(3) Work	(4) Wait Hybrid	(5) Wait Hybrid	(6) N
Total Spending	1927.22	5.19	−20.33	39.52	8.48	377
	(641.21)	(77.89)	(82.16)	(78.06)	(84.27)	
Food	1376.31	92.15	104.21	255.74	260.22	377
	(606.95)	(72.66)	(78.35)	(72.17)	(79.62)	
Voucher Staples	1061.49	18.07	22.15	285.27	282.22	377
	(571.62)	(68.52)	(73.33)	(68.05)	(72.62)	
Drinks	137.78	31.43	36.19	−37.06	−34.53	377
	(167.28)	(20.77)	(22.38)	(17.97)	(19.25)	
All Other Food	177.03	42.66	45.87	7.53	12.53	377
	(183.63)	(23.39)	(23.69)	(23.09)	(22.53)	
Non-food	550.92	−86.96	−124.54	−216.21	−251.74	377
	(519.00)	(54.57)	(53.02)	(51.39)	(54.29)	
Household Goods	156.25	−14.07	−25.67	−71.78	−80.29	377
	(241.89)	(26.40)	(26.72)	(24.15)	(26.49)	
Bath and Body	210.05	−12.72	−23.64	−26.90	−36.60	377
	(208.70)	(23.25)	(23.96)	(24.65)	(26.24)	
All Other Non-food	184.61	−60.17	−75.23	−117.53	−134.86	377
	(440.59)	(42.44)	(42.31)	(39.94)	(43.17)	
Self-reported Weekly Spending	1529.43	13.63	−81.39	−207.47	−301.55	344
	(1402.71)	(218.17)	(192.19)	(168.70)	(168.05)	
Joint test (p-value)		0.38	0.08	0.00	0.00	
Includes controls		No	Yes	No	Yes	

Notes: Results from OLS regressions comparing purchases at PBK using vouchers received in the experiment. Column 1 is the mean (and standard deviation) of the control group, which is the Wait-Unrestricted group. The columns show the effect of the Work (columns 2 and 3) and Wait-Hybrid (columns 4 and 5) treatments, relative to Wait-Unrestricted, with and without controls. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables. All spending in Kenyan Shillings.

Figure 1: Histogram of Wait-Unrestricted Staples Spending
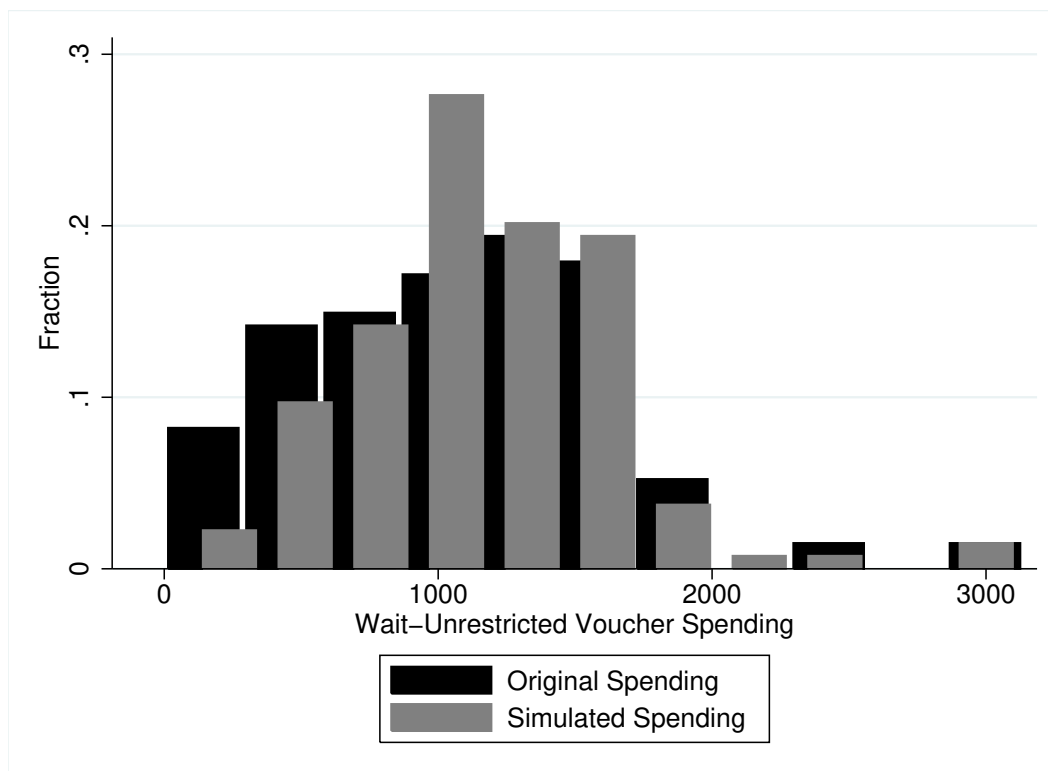
Histogram of staples spending in the Wait-Unrestricted group. The black bars represent the original staple food spending of the Wait-Unrestricted group, which has not been altered at all. The gray bars are the result of the simulated spending, where we artifically increase the amount of staples spending to be 50% of total spending for those participants who did not spend 50% of their total receipts on staple goods (for those who did spend 50% or more, their spending is unchanged). Note the rightward shift in staples spending, as we force under-spenders to move "to the kink".

spenders, in general) do rationally reallocate their budgets. Therefore, it is necessary to test the robustness of the flypaper effect conclusion. We do this by running a simulation exercise involving the mechanical creation of a "counterfactual" unrestricted group. Note that in the Wait-Hybrid group, participants were forced to spend at least half of their vouchers on staples. Therefore, in our counterfactual unrestricted group, we create a new variable where, for each participant in the Wait-Unrestricted group, staples spending is set to represent 50% of total voucher spending for all individuals for whom staples spending is actually less than 50%. In those cases where staple spending is already more than 50% of total voucher spending for a Wait-Unrestricted participant, staples spending in the counterfactual group is unchanged. Figure 1 shows the rightward shift in "staples spending" by the counterfactual group relative to the actual Wait-Unrestricted group as we artificially increase the amount of staples spending for some participants in the Wait-Unrestricted group.

We then repeat our analysis from Table 6, but use the new counterfactual unrestricted group as the comparison group instead of the actual Wait-Unrestricted group. The results are presented in Table 7. Notice the

Table 7: Robust Specification - Consumption Totals

	(1) Control Mean (SD)	(2) Wait Hybrid	(3) Wait Hybrid	(4) N
Voucher Staples (Robust)	1204.19 (465.43)	142.57 (61.71)	136.20 (65.92)	377
Includes controls		No	Yes	

Notes: Results of simulation exercise with the *counterfactual* Wait-Unrestricted treatment serving as the omitted comparison group (to create this group, anyone in Wait-Unrestricted with staples spending less than half of total spending had staples spending reweighted to represent no less than half of their total spending. All spending in Kenyan Shillings.

mean of staples spending for the new comparison group has increased from KES 1061.49 to KES 1204.19, as we artificially increase staple food spending to at least 50% of total purchases in the counterfactual unrestricted group. Likewise, the point estimate on the influence of the flypaper effect drops from KES 285.27 to KES 142.57 without controls (from KES 282.22 to KES 136.20 with controls). Crucially, the effect remains significant at the 5 percent level, with and without controls included. Given these results, we see evidence of a flypaper effect from partial voucher restriction to staple food items, representing a roughly 12% increase in spending on staple food items.

Taken together, these results contradict the predictions of consumer theory, and support a role for mental accounting in consumer decision making in the context of social welfare payments. Specifically, they suggest that pairing a transfer restricted to certain categories of spending with a transfer that is unrestricted may lead to different spending behavior than would an equivalent increase in individual income (in the form of unrestricted vouchers). Even partial restrictions in transfer payments may therefore serve to nudge individuals toward particular consumption choices without completely restricting freedom of choice for the beneficiary.

6.4 Voucher Framing: Consumption

Finally, we turn to the effect of voucher framing on consumption choice. Recall that we randomized messaging on the voucher (to encourage spending on oneself versus on one's family) in the first week, at the individual level, then switched the messaging in the second week. Table 8 shows the results of regressions comparing expenditure in the first week across subjects receiving the differently-framed vouchers. Overall, there are few compelling results, suggesting minimal impact of framing on decisions. The only sub-category of spending that was significant was bath and body products, with those receiving the Family voucher more likely to

Table 8: Family Framing Regressions - Spending

	(1) Control Mean (SD)	(2) Family Messaging	(3) Family Messaging	(4) N
Total Spending	1101.74	−2.81	37.90	376
	(465.06)	(48.63)	(49.51)	
Food	819.74	17.24	30.77	376
	(375.64)	(41.20)	(40.80)	
Voucher Staples	644.70	2.60	9.96	376
	(357.61)	(39.65)	(39.27)	
Drinks	78.89	−3.69	−0.77	376
	(99.38)	(10.19)	(10.63)	
All Other Food	96.14	18.34	21.58	376
	(119.33)	(11.87)	(12.11)	
Non-food	282.00	−20.05	7.13	376
	(351.66)	(33.75)	(35.39)	
Household Goods	76.75	−11.05	−4.74	376
	(128.54)	(12.66)	(13.21)	
Bath and Body	94.74	28.60	39.55	376
	(120.25)	(13.57)	(14.27)	
All Other Non-food	110.51	−37.61	−27.68	376
	(304.39)	(29.00)	(30.54)	
Joint test (p-value)		0.13	0.05	
Includes controls		No	Yes	

Notes: Result of an OLS regression comparing purchases at PBK using vouchers received in the experiment. This table covers vouchers redeemed in the *first week* of the experiment. Column 1 is the mean (and standard deviation) of the control group, which is the "Self" framing treatment. Columns 2 and 3 are the treatment effect of being assigned to the "Family" framing treatment with and without controls. All spending in Kenyan Shillings. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables.

spend on these products than participants receiving the Self voucher (KES 39.55, $p < 0.01$). However, the theoretical basis for this observed effect is not clear, and neither is its economic significance. Furthermore, in the endline survey, participants who receive Family vouchers first are no more likely to report purchasing items for themselves, suggesting the voucher messaging manipulation might not have been an important consideration in the decision-making process for participants.

7 Discussion

The results of our experiment offer insights into the design of welfare programs for the urban poor. Most notably, we find that working for a social welfare benefit may have greater positive effects on psychological well-being than waiting idle for a benefit of equivalent value. This provides suggestive evidence that individual utility may be impacted not only by the material welfare benefit received, but also by the way in which that benefit was obtained. As such, we posit that the social costs of funding welfare programs may be partially

offset by pairing productive workfare to the disbursement of welfare support — not only will recipients engage in work and thus *produce* something of social benefit in exchange for welfare payments, they may also attain greater individual satisfaction and utility with the welfare program. Furthermore, while this study is not a direct test of work-linked transfer programs like the Earned Income Tax Credit in the United States, or public works programs like NREGA in India, it does offer some support for the idea that work requirements in welfare may have benefits for individual wellbeing that might be missed in a simple cost-benefit analysis of these programs.

The implications of this finding, from a welfare perspective, are interesting. It seems intuitively true that people prefer not working to working, when given the choice. It is also hard to imagine workers paying for the opportunity to work, rather than to sit idle. Yet our results suggest that individuals may experience more positive emotions and greater wellbeing when working than when not working. This is difficult to reconcile with a rational model of decision-making, where leisure would be preferable to labour in the absence of compensation. However, it could fit into a model whereby people exhibit systematic biases in their assessment of future utility. Most notably, our results are supportive of a framework in which workers' predicted utility differs systematically from their experienced utility (Kahneman et al., 1997), causing them to select into an option that offers the greatest predicted utility but not the greatest experienced utility. Of course, we do not directly test this hypothesis here, as subjects were not allowed to choose their work regimen. But this is a promising area for further research.

One might also plausibly posit that our results are evidence of "idleness aversion," whereby individuals prefer doing something (even work) rather than sitting in one place doing nothing (Hsee et al., 2010). While we agree that this is one possible interpretation, we believe this does not simply "explain away" our contention that workfare might be a preferable policy to simple welfare payments. First, our daily PANAS measures of psychological wellbeing were conducted before the start of each session. Therefore the disutility of idle waiting for that day is not likely to have been captured in these reports. Second, in our experiment, people were allowed to bring reading material or other diversions with them, somewhat alleviating the boredom which might drive idleness aversion. Anecdotally, people in the wait conditions also seemed to converse a fair amount, suggesting they were not as bored as subjects in earlier studies on idleness aversion. Third, even if idleness aversion played a role, that would only support the argument that work ordeals are preferable to ordeals in the form of "waiting in line" for payment, since working at least provides some form of productive engagement that wards off the detrimental effects of idleness.

Our results also provide evidence for the applicability of mental accounting and the flypaper effect in individual-level decision making. Specifically, we found that applying restrictions to some of the vouch-

ers led to a large increase in spending on the targeted goods overall. Given the high rates of spending on these same goods in the unrestricted voucher treatment, this represents a departure from rational consumer theory predictions, based on the notion of money (or unrestricted vouchers, in this case) as fungible. Partial targeting in transfer payments through voucher restrictions may thus be an effective policy lever that does not entirely eliminate the consumer's freedom to choose, and may lead people to allocate un-earmarked portions of their welfare benefit to desired spending categories. Our results did not, however, come to the same conclusion about voucher framing, which seemed a less effective means of influencing consumer choice.

While this study does offer suggestive evidence of the positive impacts of workfare on psychological well-being, it has limitations. First, it was designed as a lab-in-the-field study and was not intended to capture general equilibrium responses, as measured by *total* consumption of a household. Indeed, because we did not complete a comprehensive general consumption module, we cannot say for certain that changes in purchase behavior reflect overall behavior, or if changes in budget shares for spending at PBK are counterbalanced by spending elsewhere, leaving total consumption shares unchanged. Of course, any counterbalancing of this sort would need to be correlated with the treatments, which while possible, seems to us to be unlikely. Second, some important questions remain unanswered. Further research is required to determine how the nature of and motivation for work (pro-social or profit) may affect happiness and consumption choices. Indeed, our experimental design is well-suited to build on work on labor economics questions of work effort and behavioral biases (as studied by Augenblick and Rabin (2015); Kremer et al. (2015); Breza et al. (2015) and others). Further, more work is needed to understand the channels through which social policy can encourage the consumption of healthy or high-calorie foods in particular. While some research has been done on curating the choice architecture of environments where food is purchased (e.g., cafeterias, stores), little has been done on examining how framing the means of payment can prompt particular spending choices. Given the low to zero marginal cost of framing interventions, further information on their effectiveness may prove useful.

References

Aker, Jenny C, "Cash or Coupons? Testing the Impacts of Cash versus Vouchers in the Democratic Republic of Congo," 2013.

Amendah, Djesika D., Steven Buigut, and Shukri Mohamed, "Coping Strategies among Urban Poor: Evidence from Nairobi, Kenya," *PLoS ONE*, 2014, *9* (1).

Anderson, Michael L, "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American statistical Association*, 2008, *103* (484).

Arkes, Hal R, Cynthia A Joyner, Mark V Pezzo, Jane Gradwohl Nash, Karen Siegel-Jacobs, and Eric Stone, "The Psychology of Windfall Gains," *Organizational Behavior and Human Decision Processes*, 1994, *59* (3), 331–347.

Augenblick, Ned and Matthew Rabin, "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," Technical Report, Working paper, Haas School of Business and Harvard University 2015.

Banerjee, Abhijit, Abhijit Vinayak Banerjee, and Esther Duflo, *Poor economics: A radical rethinking of the way to fight global poverty*, PublicAffairs, 2011.

Beshears, John and Katherine L Milkman, "Mental Accounting and Small Windfalls: Evidence from an Online Grocer," *Journal of Economic Behavior & Organization*, 2009, *71* (2), 384–394.

Bhanot, Syon, Jiyoung Han, and Chaning Jang, "Pre-Analysis Plan: Welfare, Work, and Wellbeing: Evidence from an Informal Settlement in Kenya," 2015.

Blattman, Christopher and Paul Niehaus, "Show them the money," *Foreign Affairs*, 2014, *93* (3), 117–126.

Blumkin, Tomer, Yoram Margalioth, and Efraim Sadka, "The desirability of workfare as a welfare ordeal: Revisited," 2010.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The Morale Effects of Pay Inequality," Technical Report, Working paper 2015.

Choi, James J, David Laibson, and Brigitte C Madrian, "Mental accounting in portfolio choice: Evidence from a flypaper effect," Technical Report, National Bureau of Economic Research 2007.

Cohen, Jacob, Patricia Cohen, Stephen G West, and Leona S Aiken, *Applied multiple regression/correlation analysis for the behavioral sciences*, Routledge, 2013.

Cohen, Jessica, Pascaline Dupas, and Simone G Schaner, "Price subsidies, diagnostic tests, and targeting of malaria treatment: Evidence from a randomized controlled trial," 2012.

Currie, Janet and Firouz Gahvari, "Transfers in Cash and In-Kind: Theory Meets the Data," *Journal of Economic Literature*, June 2008, *46* (2), 333–383.

Davies, Simon, Joshy Easaw, and Atanu Ghoshray, "Mental accounting and remittances: A study of rural Malawian households," *Journal of Economic Psychology*, 2009, *30* (3), 321–3334.

Devereux, Stephen, "Cash Transfers and Social Protection," in "Cash transfer activities in southern Africa" 2006.

Fiszbein, Ariel, Norbert Rüdiger Schady, and Francisco HG Ferreira, *Conditional cash transfers: reducing present and future poverty*, World Bank Publications, 2009.

Gertler, Paul, "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment," *American Economic Review*, 2004, *94* (2), 336–341.

Hidrobo, Melissa, John Hoddinott, Amber Peterman, Amy Margolies, and Vanessa Moreira, "Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador," *Journal of Development Economics*, 2014, *107*, 144–156.

Hines, James R and Richard H Thaler, "Anomalies: The flypaper effect," *The Journal of Economic Perspectives*, 1995, pp. 217–226.

Hoynes, Hilary W, Leslie McGranahan, and Diane W Schanzenbach, "SNAP and food consumption," 2014.

Hsee, Christopher K, Adelle X Yang, and Liangyan Wang, "Idleness aversion and the need for justifiable busyness," *Psychological Science*, 2010, *21* (7), 926–930.

Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin, "Back to Bentham? Explorations of experienced utility," *The quarterly journal of economics*, 1997, pp. 375–405.

Kremer, Michael, Supreet Kaur, and Sendhil Mullainathan, "Self Control at Work," *Journal of Political Economy*, 2015, *123* (6), 1227 – 1277.

Lagarde, Mylene, Andy Haines, and Natasha Palmer, "Conditional Cash Transfers for Improving Uptake of Health Interventions in Low- and Middle-Income Countries: A Systematic Review," *Journal of the American Medical Association*, 2007, *298* (16), 1900–1910.

McKenzie, David, "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 2012, *99* (2), 210–221.

Nichols, Albert L. and Richard J. Zeckhauser, "Targeting Transfers Through Restrictions on Recipients," *The American Economic Review*, 1982, pp. 372–377.

Pan, Lei and Luc Christiaensen, "Who is Vouching for the Input Voucher? Decentralized Targeting and Elite Capture in Tanzania," *World Development*, 2012, *40* (8), 1619–1633.

Pfeffer, Jeffrey, "Six dangerous myths about pay," *Harvard business review*, 1998, *76*, 108–120.

Schwartz, Barry, *Why We Work*, Simon and Schuster, 2015.

Thaler, Richard H, "Mental Accounting Matters," *Journal of Behavioral decision making*, 1999, *12* (3), 183–206.

_ , "Mental Accounting and Consumer Choice," *Marketing Science*, 2008, *27* (1), 15–25.

Watson, David, Lee A Clark, and Auke Tellegen, "Development and validation of brief measures of positive and negative affect: the PANAS scales.," *Journal of personality and social psychology*, 1988, *54* (6), 1063.