



Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo



Workfare, wellbeing and consumption: Evidence from a field experiment with Kenya's urban poor[☆]

Syon P. Bhanot^{a,b,*}, Jiyoung Han^c, Channing Jang^b

^a Swarthmore College, 500 College Ave, Swarthmore, PA 19081, United States

^b The Busara Center for Behavioral Economics, 5th Floor, Daykio Plaza, Lane, Ngong Ln, Nairobi, Kenya

^c ideas42, United States

ARTICLE INFO

Article history:

Received 30 March 2017

Revised 7 January 2018

Accepted 9 January 2018

Available online 12 January 2018

JEL classification:

C9

D1

H3

I3

J2

O1

O2

Keywords:

Workfare

Wellbeing

Consumption decisions

Mental accounting

Flypaper effect

ABSTRACT

Restrictions like work requirements and constraints on voucher transfers are often used in social welfare systems, but little empirical evidence exists on their impact on wellbeing. We conducted a 10-day randomized experiment with 432 individuals living below the poverty line in the Kawangware settlement of Nairobi, Kenya, testing two elements of social welfare design: workfare versus welfare and restricted versus unrestricted vouchers. Participants were randomly assigned to a “Work” condition, involving daily work for unrestricted vouchers, or one of two “Wait” conditions, involving daily waiting for vouchers that were either unrestricted or partially restricted to staple foods. We find that working improved psychological wellbeing relative to waiting, suggesting that the means of implementing welfare programs may have important effects on individuals beyond the impact of monetary benefit alone. Furthermore, although the restrictions were inframarginal, partially restricted vouchers crowded-in spending on staple foods, suggesting the existence of a “flypaper effect” in spending from restricted vouchers.

© 2018 Elsevier B.V. All rights reserved.

[☆] The authors wish to thank the Weiss Family Program Fund, Harvard Law School, and Harvard University, for their financial support of this project, and the staff at the Busara Center for Behavioral Economics in Nairobi, Kenya, who assisted in the implementation of the experimental protocol. Special thanks to Claudia Newman-Martin and Alexandra de Filippo for work on design and implementation, and to Jim Reisinger for research assistance with analysis. Thank you to Dan Ariely, Johannes Haushofer, Julian Jamison, Edward Miguel, Mike Norton, Jeremy Shapiro, Richard Zeckhauser, Prachi Jain, Joshua Dean, Wayne Liou, Christopher Roth, Michael Walker, the seminar participants of the LEAP Lab in Princeton University, and conference participants at SEEDEC 2016 and NEUDC 2016 for helpful comments. Special thanks to two anonymous reviewers and the editor for their helpful comments. The randomized experiment reported in this paper was registered through the American Economic Association registry prior to data analysis; the RCT ID for this study on that registry is AEARCTR-0000788. All errors are our own.

* Corresponding author at: Swarthmore College, 500 College Ave, Swarthmore, PA 19081, United States

E-mail addresses: sbhanot1@swarthmore.edu (S.P. Bhanot), jhan@ideas42.org (J. Han), channing.jang@busaracenter.org (C. Jang).

1. Introduction

It is important to understand how the structure of social welfare programs influences the wellbeing and decision-making of target populations. Indeed, policymakers need this information to design programs to maximize recipient welfare. In practice, the most popular social welfare programs tend to include some form of restriction on welfare recipients, from conditional cash transfers, voucher transfers, feeding programs (in-kind food aid), and workfare programs (or public work programs) (Honorati et al., 2015). Indeed, one can categorize these programs as varying in their levels of restriction in two key domains: (1) restrictions on the receipt of transfers (making receipt less or more conditional on the completion of certain actions by the beneficiary, as in conditional cash transfers or workfare); and (2) restrictions on the consumption of transfers (whether transfers come in more or less fungible forms – from cash payments to vouchers with varying degrees of restriction on eligible items to food aid).

Though a growing body of research has focused on the impact of welfare programs on recipients (Aker, 2017; Baird et al., 2014; Hidrobo et al., 2014; Kabeer and Waddington, 2015), there are a number of notable gaps in the literature when it comes to the effects of welfare restrictions on recipients. In particular, though both workfare and restrictive vouchers (as opposed to unrestricted vouchers) are common tools in social welfare program design,¹ there is relatively little research on the impacts of these specific restrictions on psychological wellbeing or consumption choices, whereas this has been studied more in the context of conditional and unconditional cash transfers (Haushofer and Shapiro, 2016). Furthermore, what work does exist on consumption choices under various social welfare regimes often relies on survey data regarding consumption, rather than observable data on precise items purchased (Aker, 2017; Haushofer and Shapiro, 2016; Hidrobo et al., 2014). In this paper, we explore how these two forms of restriction often used in welfare program design influence the wellbeing and spending behavior of target populations, using a lab-in-field experiment in an urban slum in Nairobi, Kenya. Critically, our lab-in-field design and partnership with a local store gives us significant control over the experimental setting, allowing us to vary the nature of the restrictions imposed, measure the psychological impacts of these restrictions, and precisely measure exact spending behavior in-store.

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 the impact of the labor supply inherent in “working for” rather than simply receiving welfare benefits without work requirements, 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 and psychological wellbeing.

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.

Our experimental design allows us to make three important contributions to existing research. First, we provide some of the first causal evidence around the effect of workfare on the wellbeing and decisions of target populations. In doing so, we contribute evidence regarding one aspect of the behavioral economics concept of “mental accounting,” namely that earned money may be treated differently than unearned money (Jakiela, 2011; Thaler, 2008). Second, while there is a large body of research on transfers in development (Blattman and Niehaus, 2014; Fiszbein et al., 2009; Gertler, 2004; Lagarde et al., 2007),

¹ For workfare, a 2015 World Bank report (Honorati et al., 2015) documented 94 developing countries with public works programs, out of 157 countries surveyed. Furthermore, these programs existed in 30 of the 33 low-income countries surveyed. Meanwhile, examples of restricted voucher programs include the use of “restricted value vouchers” in South Sudan (Yunusu et al., 2016), the use of restrictions in “e-vouchers” in Haiti (CARE, 2016), and the Healthy Start program in the United Kingdom's National Health Service (NHS, 2017).

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.

there is less experimental research on how spending restrictions built into voucher transfers might influence consumption decisions in developing contexts. That is, while there is evidence on consumption differences from cash transfers versus in-kind transfers versus vouchers targeting broad categories of spending like food (Aker, 2017; Hidrobo et al., 2014), there is little research that we are aware of that directly studies the impact of varying the restrictiveness of vouchers (i.e. vouchers for healthy food, versus food in general). Indeed, previous experimental work on transfer modalities (cash versus vouchers versus in-kind transfers) has often, by design, not held transaction costs constant across conditions – redeeming vouchers usually involves greater transaction costs than using cash or receiving in-kind food transfers. This makes it difficult to generalize previous findings to situations where restrictive versus less restrictive vouchers are being considered, specifically. Our experimental design, however, is well-suited to provide evidence regarding the effect of varying restriction levels in voucher programs in particular. Furthermore, our setup also allows us to test another aspect of mental accounting, namely the possible existence of “flypaper effects” from restricted vouchers (Hines and Thaler, 1995), whereby money provided to individuals for specific goods “sticks where it hits,” increasing overall expenditure on these goods beyond what would be predicted in the standard consumer theory model. Third, we focus not only on an objective marker of spending behavior as an outcome, but also on the understudied measure of subjective wellbeing (which is important when considering the welfare implications of these policy options).

There are two main results from the experiment. First, receiving vouchers in exchange for working had positive impacts on psychological wellbeing, relative to waiting in exchange for the same benefit. In other words, subjects in the Work treatment reported consistently higher psychological wellbeing than those in the Wait-Unrestricted treatment on nearly all measures, though the magnitudes of our results suggest caution is warranted when extrapolating our findings. Meanwhile, we did not find large effects 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 staple foods, suggesting that the voucher restrictions were inframarginal 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 from restricted vouchers, supportive of a role for mental accounting in this instance.

In addition to the academic implications, these results provide three important policy insights. First, 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. Second, if workfare generates psychological benefits for recipients, above and beyond the individual financial benefits and the social gains from the work itself, it bolsters the argument that inducing labor supply through workfare might represent a superior method for redistributing goods and money to low-income individuals (as compared to welfare schemes that do not require labor effort). Third, our lab-in-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 in both the developing and developed world.

The paper proceeds as follows. Section 2 provides background and outlines existing literature. 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

In this paper, we explore the effects of restrictions in social welfare program design in two domains: the receipt of transfers (with workfare) and the consumption of transfers (with voucher restrictions). In this section we discuss the background for each of these, to motivate our work and place it in the context of existing literature. We also discuss the two outcome measures we explore in the paper, wellbeing and consumption, and how they connect to these two forms of social welfare program restriction specifically.

2.1. Workfare restrictions on the receipts of transfers

The first type of restriction that we study in this experiment is conditionality on the receipt of the transfer. There is significant work in existing literature that directly compares unconditional cash transfers to conditional cash transfers, with the latter referring to when a transfer is restricted to individuals who undertake certain behaviors (Baird et al., 2014, 2011; Robertson et al., 2013). However, most of these studies explore conditional cash transfers that require recipients to make human capital investments in their children (education, health, nutrition, etc. Fiszbein et al., 2009). A second, very popular type of restriction on welfare receipt is public works programs, or workfare, which require labor supply of some form from individuals in exchange for transfers. Indeed, such programs are especially common in developing countries (Dutta et al., 2014; Grosh et al., 2008; Hoddinott et al., 2012; Ravallion et al., 1993). Although the recent evidence for the efficacy of these kinds of programs is generally positive (Muralidharan et al., 2017; Ravi and Engler, 2015; Rosas and Sabarwal,

2016),³ there is little evidence regarding the effects of such programs relative to other modalities for social welfare transfers. Our experiment offers some evidence in this domain, drawing on work in psychology, public economics, and behavioral economics on motivation, social program targeting, and mental accounting.

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 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., 2015](#)).

However, while ordeals can take the form of work, they do not necessarily involve labor supply. Workfare programs, on the other hand, do require labor effort, a restriction that is justified from an economic and policy perspective because it generates durable assets for the local community and decreases government expenditure ([Blumkin et al., 2010](#)). Importantly, large public works programs like India’s National Rural Employment Guarantee Act (NREGA) or Peru’s Trabajo programs do not view this work as an ordeal, but rather as a right and means of social protection – these programs offer wage employment to qualifying individuals. While the primary aim of policymakers providing this employment may be to improve the economic situations of beneficiaries and their communities, workfare may have direct impacts on the psychological wellbeing of beneficiaries as well. A growing body of literature supports this view, arguing that the intrinsic motivation, sense of higher purpose, and feeling of “usefulness” inherent in the act of working can be a source of satisfaction from work ([Pfeffer, 1998](#); [Schwartz, 2015](#)). This implies that the labor effort inherent in “workfare” may itself 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, with the important caveat that the work we require from our participants differs from that in public works programs, in that it does not involve work that directly benefits the local community (which we discuss further in [Section 3.1](#)). Furthermore, our comparison of the Work and Wait-Unrestricted treatment groups allows us to explore whether the act of working itself—not the equivalent time cost it imposes—influences wellbeing in the context of a workfare program.

Meanwhile, the topic of how consumption decisions differ under workfare versus welfare programs 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 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 purchasing decisions.

To summarize, in our experiment we vary whether or not workfare restrictions are placed on participants and measure both consumption and wellbeing as outcomes. This allows us to answer open questions regarding the impact of the labor supply aspect of workfare on spending behavior and wellbeing, which has significant implications for the efficacy and welfare impact of existing workfare programs.

2.2. Voucher restrictions on the consumption of transfers

The second type of restriction that we study in this experiment is restrictions on consumption of the transfer, with a focus on the nature of consumption restrictions in vouchers. In existing social welfare programs, the traditional method for restricting consumption in transfers has been to provide in-kind (primarily food) benefits to individuals rather than cash. This is usually used as a policy instrument to encourage a specific form of consumption, perceived to be in the best interest of the beneficiary. Again, there is a large body of existing literature on the differential impacts of in-kind transfers as opposed to cash transfers ([Gentilini, 2016](#); [Gilligan et al., 2014](#)). Furthermore, there is a small but growing body of work comparing the use of cash welfare to the use of vouchers with minor restrictions (like “food vouchers” or “education” vouchers), which provide flexibility for the recipient but still target transfers to a broad class of consumption product (see [Hidrobo et al. \(2014\)](#) and [Aker \(2017\)](#), for example). However, vouchers programs can also be much more restrictive,

³ One exception is [Beegle et al. \(2017\)](#), who show no effect of workfare on consumption, and negative spillover effects.

limiting purchases to very precise types of goods (healthy foods, education materials, etc.). Imposing such restrictions may better serve the interests of the policymaker, who may want beneficiaries to purchase specific types of items, but may not want the logistical hassle of providing in-kind transfers. However, there is very little work studying how the extent to which vouchers are restrictive might influence consumption or wellbeing, and we offer some of the first evidence in this domain.

The provision of vouchers instead of cash (or very restrictive vouchers instead of less restrictive vouchers) as a policy tool has a direct link to traditional microeconomic theory, which 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 models in behavioral economics, 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 would occur 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 account 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). Note that this same logic would apply to a less versus more restrictive voucher, namely that a flypaper effect from a more restrictive voucher would involve individuals buying more of the item targeted by the restricted voucher above and beyond what they would with the less restrictive voucher. 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., 2009), 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 toward particular choices with restrictive vouchers or in-kind transfers. For example, the U.S. Supplemental Nutrition Assistance Program (SNAP) restricts benefits to particular types of goods (Currie and Gahvari, 2008). Furthermore, there is an ongoing debate on the usefulness of restrictive 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 vouchers, 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.

While our work belongs in this broad category of research, it differs in that it looks at the effects of variations in restriction within voucher transfers specifically, rather than comparing across transfer modalities (cash versus vouchers versus in-kind) as others have done. Indeed, we offer the first experimental evidence we are aware of on the flypaper effect's relevance when using highly restrictive vouchers. For example, while Hidrobo et al. (2014) explore food consumption when receiving cash versus food vouchers versus in-kind food transfers, they do not test the use of more versus less restrictive vouchers. Furthermore, while most previous work in this area has relied on surveys to determine expenditure (Aker, 2017; Haushofer and Shapiro, 2016; Hidrobo et al., 2014), we use objective measures of in-store spending, which allows us to measure the effect of very specific restrictions in voucher transfers on actual purchasing behavior.

Beyond the consumption effects, one might posit that the extent to which vouchers are restricted to a subset of consumption items might also influence wellbeing directly. Again, a simple consumer theory model would suggest that individuals would prefer vouchers with fewer restrictions, as they enable consumers to better optimize over a broader set of consumption bundles. Furthermore, restrictions on consumption such as vouchers or in-kind transfers may lead to “welfare stigma” (Moffitt, 1983), which is costly to individuals psychologically and drives away potential recipients from applying for or receiving transfers (Manchester and Mumford, 2009; Rinehart and McGuire, 2017). Specifically, Manchester and Mumford (2009) find that psychological costs are twice as large as time costs among food stamp and WIC participants in the U.S. Our experiment allows for a direct test of this as well.

To summarize, in our experiment we directly vary the extent to which the vouchers we provide are restrictive to the participants and measure both spending and wellbeing as outcomes. By doing so, we are able to explore important questions related to the flypaper effect and consumer choice (using precise spending data), and related to the psychological effects of voucher restriction (using wellbeing data).

3. Experiment overview

3.1. Experiment design

3.1.1. 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.⁴ Participants were initially contacted by phone and asked if they were available and willing to participate in a ten-day study.⁵ 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.⁶ 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.

3.1.2. 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 beyond the expectation that the voucher payments would require some time to process. Field officers allowed participants to bring reading materials or other diversions if they wished, and did not actively stop organic chatter among the participants during their wait. 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 such that “working” and “waiting” could be directly compared. 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. We acknowledge that this simulation differs from standard social workfare programs in which beneficiaries engage in actually productive work, but this particular design was chosen for several reasons, namely: (1) as a lab-in-field experiment, the task had to be self-contained enough to be completed within an hour so as to be directly comparable to the time demands of the Wait conditions; (2) the lentil-sorting task needed to be doable by all randomly selected participants regardless of age, gender, skill set, or literacy; and (3) this design lent itself to a direct comparison between waiting and working for the same end (vouchers) rather than introducing a secondary element of explicit higher purpose in the work condition.

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

⁴ 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.

⁵ Note that the majority of subjects contacted by phone agreed to participate. A reasonable concern from the perspective of external validity is that subjects with a lower opportunity cost of time might be more likely to agree to participate in our study. We acknowledge this, but point out that this concern applies to essentially all experimental work involving recruitment of subjects.

⁶ 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 was 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.

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).

Treatment 3: Wait-Hybrid

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 staple foods only.

Shopping at PBK

After receiving their vouchers, participants could use them to purchase goods at the PBK Nonic Supermarket, 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.

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

4.1.2. 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 survey that contained the abbreviated PANAS questions on current emotions. These were completed at the start of the sessions.

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.

4.1.3. Outcomes III: attendance

Finally, for all treatment groups, attendance and timeliness to sessions were 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 numbers of days.

Of the 432 recruited participants, 360 participated in the baseline survey, while 347 participated in the endline. Note that 25 participants were surveyed at baseline but not endline, while 12 were surveyed at endline but not baseline. The 25

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%

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 (<i>p</i> -value)		0.97	0.43			0.59	0.26	

Notes: Comparison of demographic characteristics for the three treatment groups. Each row is the result of two separate regressions, where columns 2 and 3 are the coefficients of the first regression, restricted to a sample surveyed at baseline, where columns 6 and 7 are coefficients from the second regression, restricted to a sample surveyed at endline. 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. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

“attriters” were significantly younger, better educated, and more likely to be male than participants who completed both the baseline and endline surveys.⁷

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

⁷ A 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 is as “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.

5.1. Basic specification

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

$$y_i = \beta_0 + \beta_1 \text{WORK}_i + \beta_2 \text{HYBRID}_i + \delta y_{i,t=1} + \varepsilon_i \quad (1)$$

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).⁸

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:

$$y_i = \beta_0 + \beta_1 \text{WORK}_i + \beta_2 \text{HYBRID}_i + \beta_4 X_{i,t=0} + \beta_5 T_i \times X_{i,t=0} + \delta y_{i,t=1} + \varepsilon_i. \quad (2)$$

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 effect 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 wellbeing during the two-week period. In the daily data, we have outcome measures y_{it} for individual i for $t = 1, \dots, 10$, 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:

$$y_{i,t=k} = \beta_0 + \sum_{k=2}^{10} \beta_1^k (\text{WORK}_i \times [t = k]) + \sum_{k=2}^{10} \beta_2^k (\text{HYBRID}_i \times [t = k]) + \delta y_{i,t=1} + \varepsilon_{it} \quad (3)$$

where $[t = k]$ is a dummy indicator for the k^{th} 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. Workfare and vouchers: wellbeing

We begin by discussing the treatment effects of both the Work and Wait-Hybrid conditions on affective state, relative to the Wait-Unrestricted condition. Note that the results discussed here are drawn from a single regression (Eq. (3) from Section 5.2), though the results are spread over two tables for ease of comprehension.

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. Among participants that received only unrestricted vouchers, 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, $p < .01$), pride (+0.23 SD, $p < .01$), and alertness (+0.19 SD, $p < .01$) 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 suggests that the material benefit of a workfare-style program may not be the only source of utility for an individual beneficiary, but that the act of working itself might also positively influence wellbeing.

Table 4 provides similar results, but for the Wait-Hybrid treatment relative to the Wait-Unrestricted treatment. 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, $p < .10$) 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.

Finally, we turn our attention to Table 5, which presents p -values from an F-test comparing the effect size of the Work treatment to the Wait-Hybrid treatment, relative to the Wait-Unrestricted treatment.⁹ We find that the Work condition

⁸ For individuals who are missing baseline data, we follow the dummy variable adjustment procedure from Cohen et al. (2013).

⁹ This comparison was not prespecified but run on request of an anonymous reviewer.

Table 3

Work treatment panel regressions – daily affect.

	Day of treatment									Overall effect
	2	3	4	5	6	7	8	9	10	
Excited today	0.46*** (0.13)	0.61*** (0.12)	0.34*** (0.13)	0.39*** (0.14)	0.30** (0.13)	0.31** (0.13)	0.38*** (0.13)	0.46*** (0.13)	0.02 (0.15)	0.32*** (0.08)
Upset today	−0.01 (0.13)	−0.07 (0.13)	−0.38*** (0.13)	−0.12 (0.13)	−0.08 (0.14)	−0.08 (0.13)	0.05 (0.13)	−0.19 (0.13)	−0.20 (0.13)	−0.11 (0.08)
Proud today	0.38*** (0.13)	0.16 (0.13)	0.33*** (0.13)	0.43*** (0.13)	0.33*** (0.13)	0.25** (0.12)	0.21* (0.12)	0.19 (0.12)	0.07 (0.13)	0.23*** (0.08)
Alert today	0.25* (0.14)	−0.03 (0.13)	0.23* (0.13)	0.10 (0.15)	0.20 (0.13)	0.43*** (0.13)	0.35*** (0.13)	0.37*** (0.13)	0.03 (0.13)	0.19*** (0.08)
Ashamed today	−0.01 (0.15)	0.28** (0.14)	−0.14 (0.14)	0.00 (0.15)	0.06 (0.14)	−0.03 (0.15)	0.19 (0.15)	0.06 (0.14)	0.03 (0.13)	0.05 (0.08)
Happy today	0.16 (0.12)	0.10 (0.12)	0.08 (0.12)	0.14 (0.14)	−0.08 (0.11)	0.17 (0.13)	0.10 (0.13)	0.08 (0.13)	−0.25* (0.13)	0.05 (0.07)
Joint test (<i>p</i> -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. Columns 2–10 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. This is achieved through estimating equation (3) for the full sample, for each outcome row, and reporting only the coefficient and standard error for the Work condition. “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. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

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*** (0.13)	0.47*** (0.13)	0.12 (0.13)	0.18 (0.14)	0.05 (0.13)	−0.07 (0.14)	0.07 (0.13)	0.25* (0.13)	0.03 (0.15)	0.15* (0.09)
Upset today	0.04 (0.13)	0.02 (0.14)	−0.17 (0.13)	−0.07 (0.14)	0.03 (0.14)	−0.02 (0.13)	0.22 (0.13)	−0.01 (0.13)	−0.12 (0.13)	−0.01 (0.08)
Proud today	0.15 (0.13)	−0.09 (0.14)	0.19 (0.13)	0.19 (0.13)	0.25** (0.12)	−0.02 (0.12)	0.14 (0.12)	−0.08 (0.12)	−0.06 (0.13)	0.07 (0.08)
Alert today	0.11 (0.14)	−0.14 (0.14)	0.15 (0.14)	0.01 (0.15)	0.11 (0.13)	0.21 (0.13)	0.04 (0.13)	0.17 (0.14)	0.02 (0.13)	0.06 (0.08)
Ashamed today	−0.01 (0.15)	−0.19 (0.15)	−0.03 (0.14)	−0.11 (0.16)	0.16 (0.14)	0.15 (0.15)	0.12 (0.15)	−0.11 (0.15)	0.14 (0.13)	0.02 (0.09)
Happy today	0.24** (0.12)	0.11 (0.13)	0.09 (0.13)	0.20 (0.14)	−0.03 (0.11)	0.04 (0.13)	−0.01 (0.13)	−0.02 (0.14)	−0.19 (0.13)	0.04 (0.07)
Joint test (<i>p</i> -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. Columns 2–10 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. This is achieved through estimating equation (3) for the full sample, for each outcome row, and reporting only the coefficient and standard error for the Wait-Hybrid condition. “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. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

treatment effect is significantly different than the Wait-Hybrid effect for some affective states, primarily on individual days. The treatment effect on excitement is greater on three of ten days, pride on five of ten days, alertness on two of ten days, and shame on one of ten days. Considering all days together, the overall difference in treatment effects is modest, with excited and proud the only significant differences ($p < .10$). It is important to note, however, that the Work and Wait-Hybrid conditions differed in two ways - the work requirement and the voucher restriction level. So we believe that a direct comparison of the Work versus Wait-Unrestricted conditions, where the treatment effects on affect are largest, is the most appropriate way to evaluate the psychological impact of work. Based on these results, however, we conclude that while the Work treatment results in greater affective wellbeing relative to the Wait-Unrestricted treatment, the more modest effects relative to Wait-Hybrid suggest caution when stating that the results are definitive proof of the psychological benefits of working versus waiting.

Table 5*P*-values for F-tests on daily affect differences – Work versus Wait-Hybrid.

	Day of treatment									Overall effect
	2	3	4	5	6	7	8	9	10	
Excited today	0.83	0.31	0.10	0.16	0.07*	0.01***	0.02**	0.12	0.99	0.06*
Upset today	0.67	0.55	0.11	0.66	0.42	0.65	0.22	0.17	0.54	0.28
Proud today	0.08*	0.09*	0.25	0.08*	0.51	0.03**	0.59	0.03**	0.29	0.05*
Alert today	0.31	0.42	0.55	0.56	0.51	0.09*	0.02**	0.14	0.92	0.14
Ashamed today	1.00	0.00***	0.44	0.49	0.50	0.26	0.67	0.25	0.42	0.77
Happy today	0.48	0.98	0.93	0.68	0.70	0.32	0.43	0.44	0.67	0.89

Notes: Results from an F-test comparing the effect size of participants in the Work condition to the Wait-Hybrid condition, relative to the Wait-Unrestricted condition, on a number of affective state variables, with baseline controls. Columns 2–10 report the *p*-values of the difference in effect between the Work and Wait-Hybrid treatment to the Wait-Unrestricted treatment for a given day controlling for baseline measure of outcome variable at the individual level. This is achieved through estimating equation (3) for the full sample, for each outcome row, and reporting the *p*-values from the F-test of equality of coefficients for the Work versus the Wait-Hybrid condition relative to the Wait-Unrestricted condition. "Overall Effect" also reports the *p*-values from the F-test of equality of coefficients for the Work versus the Wait-Hybrid condition relative to the Wait-Unrestricted condition, but the regression pools across all days, with day dummies and clustered standard errors on the individual level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 6

Basic specification – psychological wellbeing at endline.

	(1) Control mean (SD)	(2) Work	(3) Work	(4) Wait- Hybrid	(5) Wait- Hybrid	(6) F-test (2)–(4)=0 (<i>p</i> -value)	(7) F-test (3)–(5)=0 (<i>p</i> -value)	(8) <i>N</i>
Own happiness	–0.00 (1.00)	0.17 (0.11)	0.19* (0.11)	–0.01 (0.13)	0.00 (0.13)	0.10	0.11	345
Comparative happiness	0.00 (1.00)	0.36*** (0.12)	0.41*** (0.13)	0.20 (0.14)	0.22 (0.14)	0.20	0.14	345
Makes purchase decisions	0.70 (0.46)	0.11* (0.06)	0.02 (0.05)	–0.03 (0.06)	–0.03 (0.05)	0.02**	0.38	347
Improved self-esteem	0.80 (0.40)	0.13*** (0.04)	0.14*** (0.05)	0.07 (0.05)	0.08 (0.05)	0.13	0.16	347
Happy with purchases	0.00 (1.00)	0.33*** (0.11)	0.29*** (0.09)	0.21* (0.12)	0.15 (0.13)	0.20	0.14	342
Mental health	0.00 (1.00)	0.31*** (0.12)	0.29** (0.11)	0.07 (0.14)	0.08 (0.14)	0.06*	0.10*	346
Life prospects	–0.00 (1.00)	0.38*** (0.11)	0.40*** (0.12)	0.38*** (0.12)	0.43*** (0.12)	0.99	0.76	343
Life security (women only)	0.00 (1.00)	0.42*** (0.15)	0.47*** (0.15)	0.63*** (0.16)	0.75*** (0.16)	0.15	0.06*	258
Wellbeing index (weighted avg.)	–0.00 (1.00)	0.67*** (0.11)	0.60*** (0.11)	0.29** (0.13)	0.32** (0.13)	0.00***	0.02**	347
Joint test (<i>p</i> -value)		0.00***	0.00***	0.00***	0.00***			
Includes controls		No	Yes	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. Specifically we estimate equation (1) for columns 2 and 4 for each row, and separately estimate equation (1) for columns 4 and 6 for each row, then report the coefficients and standard errors for each. Columns 6 and 7 are the results of an F-test comparing the coefficients for Work and Wait-Hybrid within a single regression. The bottom row reports *p*-values from seemingly unrelated regressions run across all outcome variables. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Next, we review Table 6, which presents the results of OLS regressions estimating the impact of the treatment arms on various psychological wellbeing measures at endline.¹⁰ Note that the Wait-Unrestricted treatment again 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, marital 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.

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 sig-

¹⁰ In this paper, we present results of all prespecified outcomes with the exception of the Ways of Coping index. Because we failed to prespecify the way in which we would aggregate the outcomes into an index, we investigate two indices and present them in the online appendix.

Table 7

Pooled specification – psychological wellbeing at endline.

	(1) Control mean (SD)	(2) Work	(3) Work	(4) N
Own happiness	−0.00 (0.99)	0.18** (0.09)	0.19** (0.10)	345
Comparative happiness	0.09 (1.03)	0.27*** (0.10)	0.30*** (0.11)	345
Makes purchase decisions	0.69 (0.46)	0.12** (0.05)	0.03 (0.04)	347
Improved self-esteem	0.83 (0.37)	0.10*** (0.03)	0.10*** (0.04)	347
Happy with purchases	0.10 (0.94)	0.24*** (0.08)	0.22*** (0.07)	342
Mental health	0.03 (1.04)	0.28*** (0.10)	0.25** (0.10)	346
Life prospects	0.18 (0.91)	0.20** (0.09)	0.18** (0.09)	343
Life security (women only)	0.32 (1.05)	0.10 (0.13)	0.08 (0.13)	258
Wellbeing index (weighted avg.)	0.14 (1.02)	0.53*** (0.10)	0.45*** (0.09)	347
Joint test (<i>p</i> -value)		0.00***	0.00***	
Includes controls		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 pooled control group, which is the Wait-Unrestricted and Wait-Hybrid groups combined. 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 treatment (columns 2 and 3), relative to the pooled Wait-Unrestricted and Wait-Hybrid groups, with and without controls. The bottom row reports *p*-values from seemingly unrelated regressions run across all outcome variables. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

nificantly higher levels of psychological wellbeing at endline than participants in the Wait-Unrestricted treatment on nearly all measures, as seen in Table 6. To test for overall significance, we create an index of the component measures based on (Anderson, 2008), reported in Table 6 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 ($p < .01$). The significance of the effect holds both with and without controls.

Notably, relative to the Wait-Unrestricted group, participants in the Work treatment self-reported better mental health (+0.29 SD, $p < .05$) and improved self-esteem (+0.14 SD, $p < .01$) at endline. Those in the Work treatment also show large increases in comparative happiness, happiness with their purchases, life prospects, and life security (women only), relative to the Wait-Unrestricted group ($p < .01$ for all). We see a smaller effect on self-reported own-happiness (+0.19 SD, $p < .10$) at endline, though we see a null-effect on incidental daily happiness from the daily analysis in Table 3.

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 6 shows that participants in the Wait-Hybrid treatment reported greater life security (women only, +0.75 SD, $p < .01$) and life prospects (+0.43 SD, $p < .01$) at endline than those in the Wait-Unrestricted treatment.

In light of the Wait-Hybrid treatment’s significant impacts, we pooled all participants in the Wait conditions and compared their psychological wellbeing measures with those in the Work treatment as a robustness check. In Table 7, we see that the effects of the Work treatment on endline measures of happiness with purchases, self-esteem, general happiness, mental health, and life prospects hold when compared to the combined Wait conditions—albeit with smaller magnitudes.

While the Work treatment retains its effects compared to the Pooled comparison group in Table 7, the significance of the coefficients for both the Wait-Hybrid and Work conditions relative to the Wait-Unrestricted condition suggest caution is required when claiming that working was the reason for the endline increases in self-reported life security, life prospects, and household decision-making. Contrary to the literature, we find that restrictions on consumption through voucher restrictions do not lead to “welfare stigma.” However, this is likely due to the nature of the experiment, where the comparison “Wait-Unrestricted” group also uses vouchers, meaning that if negative psychological impacts flow through restrictions on vouchers versus cash rather than restrictions on what vouchers can be redeemed for, they would not be picked up in this experiment. Furthermore, the Wait-Hybrid condition may have influenced positive wellbeing measures through a different channel (for example, receiving restricted vouchers that guaranteed access to key staple goods may have empowered women in the household by binding the hands of male influence, though this is entirely speculative). Overall, however, a direct comparison of the wellbeing effects from Work versus Wait-Hybrid suggests that the increases in wellbeing from working outweigh those from voucher restrictions (see the coefficients on the Wellbeing Indices for the two conditions Table 6 – 0.60 SD for Work versus 0.32 SD for Wait-Hybrid, with $p < .05$ for the difference).

Table 8

Basic specification – consumption totals.

	(1) Control mean (SD)	(2) Work	(3) Work	(4) Wait- Hybrid	(5) Wait- Hybrid	(6) F-test (2)–(4)=0 (p-value)	(7) F-test (3)–(5)=0 (p-value)	(8) N
Total spending	1927.22 (641.21)	5.19 (77.89)	–20.33 (82.16)	39.52 (78.06)	8.48 (84.27)	0.66	0.72	377
Food	1376.31 (606.95)	92.15 (72.66)	104.21 (78.35)	255.74*** (72.17)	260.22*** (79.62)	0.02**	0.03**	377
Voucher staples	1061.49 (571.62)	18.07 (68.52)	22.15 (73.33)	285.27*** (68.05)	282.22*** (72.62)	0.00***	0.00***	377
Drinks	137.78 (167.28)	31.43 (20.77)	36.19 (22.38)	–37.06** (17.97)	–34.53* (19.25)	0.00***	0.00***	377
All other food	177.03 (183.63)	42.66* (23.39)	45.87* (23.69)	7.53 (23.09)	12.53 (22.53)	0.14	0.18	377
Non-food	550.92 (519.00)	–86.96 (54.57)	–124.54** (53.02)	–216.21*** (51.39)	–251.74*** (54.29)	0.00***	0.00***	377
Household goods	156.25 (241.89)	–14.07 (26.40)	–25.67 (26.72)	–71.78*** (24.15)	–80.29*** (26.49)	0.00***	0.01**	377
Bath and body	210.05 (208.70)	–12.72 (23.25)	–23.64 (23.96)	–26.90 (24.65)	–36.60 (26.24)	0.53	0.58	377
All other non-food	184.61 (440.59)	–60.17 (42.44)	–75.23* (42.31)	–117.53*** (39.94)	–134.86*** (43.17)	0.01**	0.03**	377
Self-reported weekly spending	1529.43 (1402.71)	13.63 (218.17)	–81.39 (192.19)	–207.47 (168.70)	–301.55* (168.05)	0.29	0.28	344
Joint test (p-value)		0.38	0.08*	0.00***	0.00***			
Includes controls		No	Yes	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. Specifically we estimate equation (1) for columns 2 and 4 for each row, and separately estimate equation (1) for columns 4 and 6 for each row, then report the coefficients and standard errors for each. Columns 6 and 7 are the results of an F-test comparing the coefficients for Work and Wait-Hybrid within a single regression. The bottom row reports p-values from seemingly unrelated regressions run across all outcome variables. All spending is in Kenyan Shillings. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

6.2. Workfare and vouchers: consumption

Table 8 shows the results of OLS regressions that compare the relationship between the treatment arms and both specific and broad categories of grocery purchases. To test the hypotheses around mental accounting and the flypaper effect in more versus less restrictive vouchers, we focus on differences in spending between the Wait-Hybrid and Wait-Unrestricted groups (the latter is the omitted comparison group in Table 8). 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 ($p < .01$). 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, $p < .01$). So it is not the case that the greater staple 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 led to greater food spending overall. This suggests that targeted vouchers may be a useful 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 staples spending in the Wait-Unrestricted treatment. As column 1 in Table 8 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 spending (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 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

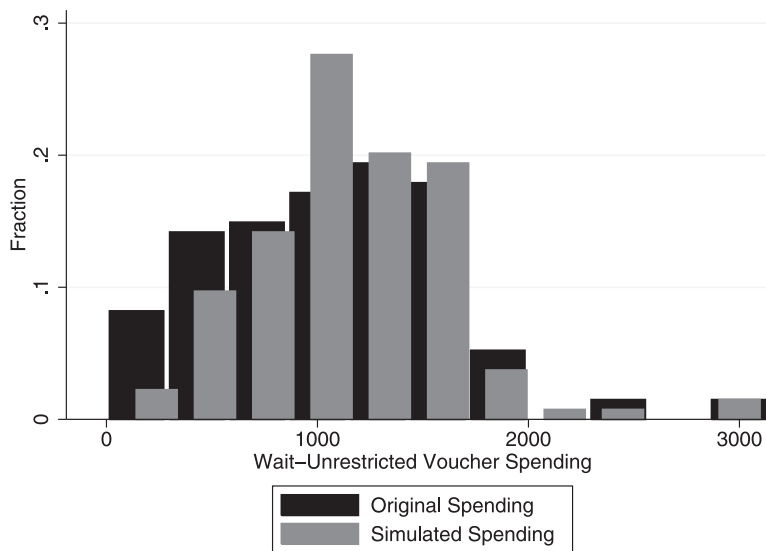


Fig. 1. Histogram of staples spending in the Wait-Unrestricted group. The black bars show the distribution of the original staple food spending in the Wait-Unrestricted group, which has not been altered at all. The gray bars are the result of the simulated spending exercise, where we artificially increase 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”.

Table 9

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). The equation estimated is an OLS regression estimated twice, with and without controls. All spending is in Kenyan Shillings. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

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 staples spending is already more than 50% of total voucher spending for a Wait-Unrestricted participant, staples spending in the counterfactual group is unchanged. Fig. 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 8, but use the new counterfactual unrestricted group as the comparison group instead of the actual Wait-Unrestricted group. The results are presented in Table 9. Notice the 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 (and from KES 282.22 to KES 136.20 with controls). Crucially, the effect remains significant at the 5% level, with and without controls included. Given these results, we see our study as providing evidence of a flypaper effect from partial voucher restriction to staple food items, representing a roughly 12% increase in spending on staple food items in our sample.

Finally, we compare spending in the Work and Wait-Unrestricted conditions directly. The results suggest minimal differences in specific categories of consumption spending between these conditions. The point estimates do provide some weak evidence that participants in the Work condition purchased more non-staple, non-drink food items (KES 45.87, $p < .10$) and less non-food (KES -124.54, $p < .05$) 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 .08, meaning we cannot confidently assert that spending 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).

To summarize, the comparative effects of the Wait-Hybrid treatment compared to the Wait-Unrestricted treatment somewhat contradict the predictions of consumer theory, and support a role for mental accounting in consumer decision-making in the context of voucher restrictions. Specifically, these results suggest that vouchers with tighter consumption restrictions may lead to different spending behavior than would unrestricted vouchers, even if the voucher restrictions are inframarginal. 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.

7. Discussion

The results of our experiment offer insights into the design of welfare programs for the urban poor. Notably, we find that working for a social welfare benefit may have greater positive effects on psychological wellbeing than waiting for a benefit of equivalent value. This provides suggestive evidence that individual utility may be impacted not only by the material welfare benefit received (restrictions on the consumption of transfers), but also by the way in which that benefit was obtained (restrictions on the receipt of transfers). As such, we posit that the social costs of funding welfare programs may be partially offset by pairing work requirements with the disbursement of welfare support – not only will recipients engage in work and potentially *produce* something of social benefit in exchange for welfare payments, they may also obtain utility from the act of supplying labor in the workfare program (and not just from the material reward from transfers). Note that this is the case in our experiment even though the work we had participants do was devoid of a social purpose—perhaps these effects would be even larger from work imbued with greater meaning.

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 people paying for the opportunity to work, rather than sitting in a waiting room (especially when diversions like conversation or reading are allowed in the latter). 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 labor in the absence of compensation. However, it could fit into a model whereby people obtain psychic benefits from the sense of accomplishment deriving from task completion during labor supply (Pfeffer, 1998; Schwartz, 2015), but they do not fully appreciate these benefits *ex-ante* (and only learn to appreciate them with experience). Of course, this interpretation is speculation. However, our work does at least suggest that this is a promising area for further research, particularly given the increased use of workfare and public works programs worldwide.

One might argue 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, in our experiment, people were allowed to bring reading material or other diversions with them, somewhat alleviating the boredom that 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. Second, 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 (though we cannot rule out the possibility that they might capture disutility from anticipated boredom). 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 vouchers 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.

Importantly, we do not see our results as contradicting other work that does not find differential spending behavior with cash transfers versus voucher transfers in the developing world (like Hidrobo et al., 2014 and Aker, 2017 for example). Indeed, we are testing different modalities than Hidrobo et al. (2014) and Aker (2017) – we explore restrictions within the category of voucher transfers, whereas they test cash transfers versus food vouchers of a similar size.¹¹ Therefore, our results speak more to the design of voucher programs specifically, rather than comparing across the broader categories of welfare transfer programs explored in previous work.

While this study does offer suggestive evidence of the positive impacts of workfare on psychological wellbeing, 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

¹¹ In addition, the context around the transfer in our experiment is also quite different than that in previous work. For example, the transfers studied by Hidrobo et al. (2014) were part of a broader WFP program targeting food security and nutritious eating in particular.

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); Breza et al. (Forthcoming); Kremer 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 the means of payment can prompt particular spending choices.

References

- Aker, J., 2017. Comparing cash and voucher transfers in a humanitarian context: evidence from the democratic republic of Congo. *World Bank Econ. Rev.* 31 (1), 44–70.
- Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *J. Am. Stat. Assoc.* 103 (484), 1481–1495.
- Arkes, H.R., Joyner, C.A., Pezzo, M.V., Nash, J.G., Siegel-Jacobs, K., Stone, E., 1994. The psychology of windfall gains. *Organ. Behav. Hum. Decis. Process.* 59 (3), 331–347.
- Augenblick, N., Rabin, M., 2015. An experiment on time preference and misprediction in unpleasant tasks. Working paper. Haas School of Business and Harvard University.
- Baird, S., Ferreira, F.H., Özler, B., Woolcock, M., 2014. Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *J. Dev. Eff.* 6 (1), 1–43.
- Baird, S., McIntosh, C., Özler, B., 2011. Cash or condition? Evidence from a cash transfer experiment. *Q. J. Econ.* 126 (4), 1709–1753.
- Beegle, K., Galasso, E., Goldberg, J., 2017. Direct and indirect effects of Malawi's public works program on food security. *J. Dev. Econ.* 128, 1–23.
- Beshears, J., Milkman, K.L., 2009. Mental accounting and small windfalls: evidence from an online grocer. *J. Econ. Behav. Organ.* 71 (2), 384–394.
- Bhanot, S., Han, J., Jang, C., 2015. Pre-analysis plan: welfare, work, and wellbeing: evidence from an informal settlement in Kenya, AEA RCT Registry (RCT ID: AEARCTR-0000788).
- Blattman, C., Niehaus, P., 2014. Show them the money. *Foreign Aff.* 93 (3), 117–126.
- Blumkin, T., Margalioth, Y., Sadka, E., 2010. The desirability of workfare as a welfare ordeal: revisited, Discussion Paper No. 5130, Institute for the Study of Labor (IZA).
- Breza, E., Kaur, S., Shamasani Y., The morale effects of pay inequality, *The Quarterly Journal of Economics*. Forthcoming.
- CARE, 2016. Tool overview: March 2016: E-Voucher Transfer Platform. Technical Report. CARE Haiti.
- Choi, J.J., Laibson, D., Madrian, B.C., 2009. Mental accounting in portfolio choice: evidence from a flypaper effect. *American Economic Review* 99 (5), 2085–2095.
- Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2013. Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.
- Cohen, J., Dupas, P., Schaner, S.G., 2015. Price subsidies, diagnostic tests, and targeting of malaria treatment: evidence from a randomized controlled trial. *Am. Econ. Rev.* 105 (2), 609–645.
- Currie, J., Gahvari, F., 2008. Transfers in cash and in-kind: theory meets the data. *J. Econ. Lit.* 46 (2), 333–383.
- Davies, S., Easaw, J., Ghoshray, A., 2009. Mental accounting and remittances: a study of rural Malawian households. *J. Econ. Psychol.* 30 (3), 321–3334.
- Devereux, S., 2006. Cash transfers and social protection. Cash transfer activities in Southern Africa https://sarpn.org/documents/d0002199/Cash_Transfer_Devereux_Oct2006.pdf.
- Dutta, P., Murgai, R., Ravallion, M., Van de Walle, D., 2014. Right to work? Assessing India's employment guarantee scheme in Bihar. World Bank Publications.
- Fiszbein, A., Schady, N., Ferreira, F. H., Grosh, M., Kelleher, N., Olinto, P., Skoufias, E., 2009. Conditional cash transfers: reducing present and future poverty, World Bank Publications.
- Gentilini, U., 2016. The revival of the cash versus food debate: new evidence for an old quandary. Policy Research Working Paper Series 7584. The World Bank.
- Gertler, P., 2004. Do conditional cash transfers improve child health? Evidence from Progresa's control randomized experiment. *Am. Econ. Rev.* 94 (2), 336–341.
- Gilligan, D., Hidrobo, M., Hoddinott, J., Roy, S., Schwab, B., 2014. Much ado about modalities: multicountry experiments on the effects of cash and food transfers on consumption patterns. Agricultural and Applied Economics Association, Minneapolis, Minnesota.
- Grosh, M., Del Ninno, C., Tesliuc, E., Ouerghi, A., 2008. For protection and promotion: the design and implementation of effective safety nets. World Bank Publications.
- Haushofer, J., Shapiro, J., 2016. The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *Q. J. Econ.* 131 (4), 1973–2042.
- Hidrobo, M., Hoddinott, J., Peterman, A., Margolies, A., Moreira, V., 2014. Cash, food, or vouchers? Evidence from a randomized experiment in Northern Ecuador. *J. Dev. Econ.* 107, 144–156.
- Hines, J.R., Thaler, R.H., 1995. Anomalies: the flypaper effect. *J. Econ. Perspect.* 217–226.
- Hoddinott, J., Berhane, G., Gilligan, D.O., Kumar, N., Seyoum Taffesse, A., 2012. The impact of Ethiopia's productive safety net programme and related transfers on agricultural productivity. *J. Afr. Econ.* 21 (5), 761–786.
- Honorati, M., Gentilini, U., Yemtsov, R.G., 2015. The state of social safety nets 2015. World Bank. <http://documents.worldbank.org/curated/en/2015/07/24741765/state-social-safety-nets-2015> (accessed 16.02.15)
- Hsee, C.K., Yang, A.X., Wang, L., 2010. Idleness aversion and the need for justifiable busyness. *Psychol. Sci.* 21 (7), 926–930.
- Jakiela, P., 2011. Social preferences and fairness norms as informal institutions: experimental evidence. *Am. Econ. Rev.* 101 (3), 509–513.
- Kabeer, N., Waddington, H., 2015. Economic impacts of conditional cash transfer programmes: a systematic review and meta-analysis. *J. Dev. Eff.* 7 (3), 290–303.
- Kremer, M., Kaur, S., Mullainathan, S., 2015. Self control at work. *J. Polit. Econ.* 123 (6), 1227–1277.
- Lagarde, M., Haines, A., Palmer, N., 2007. Conditional cash transfers for improving uptake of health interventions in low- and middle-income countries: a systematic review. *J. Am. Med. Assoc.* 298 (16), 1900–1910.
- Manchester, C. F., Mumford, K. J., 2009. How costly is welfare stigma? Separating psychological costs from time costs, Social Science Research Network.
- McKenzie, D., 2012. Beyond baseline and follow-up: the case for more t in experiments. *J. Dev. Econ.* 99 (2), 210–221.
- Moffitt, R., 1983. An economic model of welfare stigma. *Am. Econ. Rev.* 73 (5), 1023–1035.
- Muralidharan, K., Niehaus, P., Sukhtankar, S., 2017. General equilibrium effects of (improving) public employment programs: experimental evidence from India. NBER (Working Paper No. 23838).
- NHS, 2017. The United Kingdom national health service: healthy start. NHS <https://www.healthystart.nhs.uk/>. (accessed 17.12.01).
- Nichols, A.L., Zeckhauser, R.J., 1982. Targeting transfers through restrictions on recipients. *Am. Econ. Rev.* 72 (2), 372–377.

- Pan, L., Christiaensen, L., 2012. Who is vouching for the input voucher? Decentralized targeting and elite capture in Tanzania. *World Dev.* 40 (8), 1619–1633.
- Pfeffer, J., 1998. Six dangerous myths about pay. *Harv. Bus. Rev.* 76, 108–120.
- Ravallion, M., Datt, G., Chaudhuri, S., 1993. Does Maharashtra's employment guarantee scheme guarantee employment? Effects of the 1988 wage increase. *Econ. Dev. Cult. Chang.* 41 (2), 251–275.
- Ravi, S., Engler, M., 2015. Workfare as an effective way to fight poverty: the case of India's NREGS. *World Dev.* 67, 57–71.
- Rinehart, C.S., McGuire, J.W., 2017. Obstacles to takeup: Ecuador's conditional cash transfer program, the Bono de Desarrollo Humano. *World Dev.* 97, 165–177.
- Robertson, L., Mushati, P., Eaton, J., Dumba, L., Mavise, G., Makoni, J., Schumacher, C., Crea, T., Monasch, R., Sherr, L., Garnett, G., Nyamukapa, C., Gregson, S., 2013. Effects of unconditional and conditional cash transfers on child health and development in Zimbabwe: a cluster-randomised trial. *Lancet* 381 (9874), 1283–1292.
- Rosas, N., Sabarwal, S., 2016. Can you work it? Evidence on the productive potential of public works from a youth employment program in Sierra Leone. World Bank Policy Research Working Paper. 7580
- Schwartz, B., 2015. Why we work. Simon and Schuster.
- Thaler, R.H., 1999. Mental accounting matters. *J. Behav. Decis. Mak.* 12 (3), 183–206.
- Thaler, R.H., 2008. Mental accounting and consumer choice. *Mark. Sci.* 27 (1), 15–25.
- Watson, D., Clark, L., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Personal. Soc. Psychol.* 54 (6), 1063–1070.
- Yunusu, E., Sibanda, G., Markham, M., 2016. Cash-based programming to address hunger in conflict-affected South Sudan: a case study. Technical Report. World Vision South Sudan.