American Economic Review 2017, 107(4): 1165–1206
https://doi.org/10.1257/aer.20150503

Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia[†]

By Christopher Blattman, Julian C. Jamison, and Margaret Sheridan*

We show that a number of noncognitive skills and preferences, including patience and identity, are malleable in adults, and that investments in them reduce crime and violence. We recruited criminally engaged men and randomized one-half to eight weeks of cognitive behavioral therapy designed to foster self-regulation, patience, and a noncriminal identity and lifestyle. We also randomized $200 grants. Cash alone and therapy alone initially reduced crime and violence, but effects dissipated over time. When cash followed therapy, crime and violence decreased dramatically for at least a year. We hypothesize that cash reinforced therapy's impacts by prolonging learning-by-doing, lifestyle changes, and self-investment. (*JEL* D12, D83, H23, I32, K42, O15, O17)

In many countries, poor young men exhibit high rates of violence, crime, and other *antisocial* behaviors. In addition to their direct costs, crime and instability hinder economic growth by reducing investment or diverting productive resources to security. In fragile states, such men are also targets for mobilization into election intimidation, rioting, and rebellion.[1]

* Blattman: Pearson Institute for the Study and Resolution of Global Conflicts and Harris School of Public Policy, University of Chicago, 1155 E. 60th Street, Chicago, IL 60637 (e-mail: blattman@uchicago.edu); Jamison: The World Bank Global Insights Initiative, MC4-404, 1818 H Street NW, Washington, DC 20433 (e-mail: julison@gmail.com); Sheridan: Clinical Psychology, University of North Carolina at Chapel Hill, 235 E. Cameron Street, Office 248, Chapel Hill, NC 27599 (e-mail: sheridan.margaret@unc.edu). For implementation we thank Global Communities and the Network for Empowerment and Progressive Initiatives (NEPI). For comments we thank Thomas Abt, Jeannie Annan, Alex Coppock, Patricio Dalton, Stefano DellaVigna, Ruben Enikolopov, Roland Fryer, Don Green, Jonas Hjort, Macartan Humphreys, Dean Karlan, Larry Katz, Shamus Khan, Winston Lin, Jens Ludwig, Mattias Lundberg, Sendhil Mullainathan, Anandi Mani, Chris Muller, Suresh Naidu, Jonathan Pinckney, Vincent Pons, Nancy Qian, Steve Radelet, Gautam Rao, Adam Reich, Alix Rule, Cyrus Samii, Rachel Strohm, Francesco Trebbi, Eric Verhoogen, the referees, and participants at numerous conferences and seminars. This study was funded by the National Science Foundation (SES-1317506), the World Bank's Learning on Gender and Conflict in Africa (LOGiCA) trust fund, the World Bank's Italian Children and Youth (CHYAO) trust fund, the UK Department for International Development (DFID) via the Institute for the Study of Labor (IZA), a Vanguard Charitable Trust, the American People through the United States Agency for International Developments (USAID) DCHA/CMM office, and the Robert Wood Johnson Health and Society Scholars Program at Harvard University (Cohort 5). The contents of this study are the responsibility of authors and do not necessarily reflect the views of their employers, funding agencies, or governments. Emma Tsui trained our qualitative researchers. Philip Blue, Natalie Carlson, Samantha DeMartino, Camelia Dureng, Mathilde Emeriau, Yuequan Guo, Brittany Hill, Tricia Koroknay-Palicz, Rebecca Littman, Ryan Luby, Ben Morse, Richard Peck, Patryk Perkowski, Colombine Peze-Heidsieck, Katherine Rodrigues, Carmel Salhi, Joe St. Clair, Helen Smith, Gwendolyn Taylor, Abel Welwean, Prince Williams, Xing Xia, Adam Xu, and John Zayzay provided research assistance through Innovations for Poverty Action. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

[†] Go to https://doi.org/10.1257/aer.20150503 to visit the article page for additional materials and author disclosure statement(s).

[1] For example, poor urban young men were recently recruited into election violence in Sierra Leone (Christensen and Utas 2008) and as mercenaries in Côte d'Ivoire (Blattman and Annan 2016).

Two of the most common government responses are policing and job creation. Both take the person as they are and try to change their incentives or simply incarcerate them (Becker 1968; Draca and Machin 2015). This paper investigates an alternative: rehabilitation, or changing behavior by shaping people's underlying skills and preferences.

A large literature has shown that a broad set of so-called *noncognitive* skills, especially self-control, predict long-run economic performance and criminal activity.[2] These skills respond to investment, especially in childhood (Cunha, Heckman, and Schennach 2010). They are fostered by family, schools, and communities. There is little evidence, however, on the returns to late-stage noncognitive investments, and so it is unclear whether by adulthood self-investment or interventions can shape noncognitive skills and hence behavior (Heckman and Kautz 2014; Hill et al. 2011). It is also unclear what specific skills are both important and malleable.

To investigate, we recruited 999 of the highest-risk men in Liberia's capital, generally aged 18 to 35. Most were engaged in part-time theft and drug dealing, and regularly had violent confrontations with each other, community members, and police.

We experimentally ran two interventions. One was an eight-week program of group cognitive behavior therapy (CBT) called the STYL program (Sustainable Transformation of Youth in Liberia). We assigned offers by lottery. Following this, we held a second lottery for a $200 grant, about three months' wages. The cash was partly a measurement tool, to see if therapy affected economic decisions. The cash was also a treatment, in the sense that it could stimulate legal self-employment, and we included it to compare therapy to a rise in the returns to legal work.[3] Experimentally, subjects either received offers of therapy alone, cash, therapy then cash, or neither. To deliver both treatments cost about $530 per person.

CBT is a therapeutic approach that can be used to treat a range of harmful beliefs and behaviors, including depression, anger, and impulsivity. First, CBT tries to make people aware of and challenge harmful, automatic patterns of thinking or behavior. Second, it tries to disrupt these patterns of thinking and to foster better ones by having people practice new skills and behaviors—learning by doing. A Liberian nonprofit, the Network for Empowerment and Progressive Initiatives (NEPI), designed and ran STYL. NEPI facilitators were themselves ex-combatants or ex-criminals who graduated from prior NEPI programs.

Among noncognitive skills, NEPI designed STYL to focus on forward-looking behavior and self-control. By self-control, psychologists and criminologists typically mean one's short-term abilities to regulate emotions and to be resistant to impulse, as well as more sustained abilities to be planful, persevering, and patient. This concept has parallels to economic time preferences, and we measure them in the manner of both fields. Becoming more self-controlled and forward-looking are central components of many programs, from preschool to rehabilitation therapy.[4] The curriculum focused on helping men foster skills of planning, goal-setting, reflection, deliberate decision-making, and controlling emotions and impulses.

[2] For example, Nagin and Pogarsky (2004); Heckman, Stixrud, and Urzua (2006); Borghans et al. (2008).

[3] Evidence from East Africa suggests that the poor and unemployed are credit-constrained and have high returns to cash (Haushofer and Shapiro 2016; Blattman, Fiala, and Martinez 2014; Blattman et al. 2016).

[4] For example, Gottfredson and Hirschi (1990) and Borghans et al. (2008). As an example of an intervention, the famous Perry Preschool Program emphasized the ability of young children to plan tasks, to execute their plans, and to review their work in social groups (Almlund et al. 2011).

The therapy also encouraged nonviolent, noncriminal behavior and lifestyles by fostering a change in the men's social identity. A premise of STYL was that the men self-identified as outcasts and did not hold themselves to the standards of mainstream society. The therapy tried to persuade the men that they could change who they were and how they were perceived. NEPI facilitators modeled this identity change. They walked the men through basic steps, such as changing their appearance, engaging in normal social interactions, and behaving more cooperatively. They discouraged drug use and association with bad peers. Therapy also required men to practice going to supermarkets, banks, and other "normal" places.

Research in both psychology and economics supports the idea that social identity and associated values influence behavior, and that both can change. This literature treats values as direct utility benefits or penalties from acting in accordance with or against a set of preferences (Bénabou and Tirole 2004; Almlund et al. 2011). Akerlof and Kranton (2000) and Jolls, Sunstein, and Thaler (1998) both argue that these preferences or values are tied to a person's social identity, and that to some extent people can change their perceived social category and with it values that reward and penalize certain behaviors.

There are striking parallels between STYL and socialization into militaries, street culture, gangs, and armed groups. Such groups use similar techniques (appearance change, practice, modeling) to shape young men's social identity and behavior (Vigil 2003; Wood 2008; Maruna and Roy 2007). NEPI designed STYL to reverse this process.

We surveyed the men beforehand, a few weeks after the interventions, and a year later. Most had no fixed address, phone, or even name. Despite this mobility, we reinterviewed 92.4 percent. We rely on self-reported data since (like most poor and fragile states) there are no administrative data. We did not necessarily trust self-reports, and so we attempted to recheck and validate behaviors such as stealing through in-depth interviews with a subsample.

We approached roughly 1,500 high-risk men, and 999 agreed to enter the study. Of those assigned to therapy, nearly all attended at least one day, and two-thirds completed it. The higher-risk men were the most likely to finish.

Men who received therapy reduced their antisocial behaviors dramatically, roughly 0.2 standard deviations compared to the control group. Within a few weeks of therapy, for example, we observe large reductions in an index of behaviors, including stealing and drug selling. With therapy alone, these effects diminished after a year. When therapy was followed by cash, however, the reductions in an index of all antisocial behaviors were lasting.

The therapy probably worked through many channels, and we see evidence of improvement in two of the hypothesized channels: time preferences and noncriminal identity/values, with time preference changes most persistent. There is also some evidence of improvements in positive self-regard, plus temporary changes in drug abuse and noncriminal social networks. With therapy alone, a broad index of all these intermediary outcomes changes diminished after a year, just as we saw with antisocial behavior itself. When therapy was followed by cash, however, the overall changes in these intermediary outcomes were lasting and fairly large: at least 0.25 standard deviations. If we account for multiple comparisons, it becomes difficult to single out any one mechanism, but the largest and most statistically significant change is in time preferences.

How was cash used? Regardless of therapy, little of the grant was spent on drugs or "wasteful" things. Most funds were invested in business or saved. Cash led to a short-term increase in an index of economic performance (including income, savings, employment, and investment), due largely to increased petty trading. After a year, however, these gains disappeared, partly because most men were robbed regularly, irrespective of treatment.

The fact that the grant was crucial to sustaining therapy's effects is our most unexpected and important finding. As we find no sustained effect of cash on earnings, cash clearly did not raise the opportunity cost of antisocial behavior after a year. Thus, economic performance does not explain the sustained effect of therapy plus cash on crime and aggression. Drawing on qualitative interviews and psychological theory, we suggest that the short-term increase in income and legal employment helped to solidify therapy's impact on noncognitive skills and preferences. Specifically, for a few months after therapy, cash allowed men to project a new self, to stave off homelessness and stealing, and practice the self-control and future orientation started by therapy. This hypothesis will be important to test in future research.

An obvious concern is our reliance on self-reported data. We argue that misreporting is unlikely to drive our results for two reasons. The first is the pattern of treatment effects: 12–13-month impacts from therapy plus cash, but not from cash or therapy only. Systematic measurement error would need to be correlated with the "both" treatment arm only. This seems feasible but unlikely, especially given the magnitudes of the impacts. To check further, we attempted to validate a subset of questions using intensive qualitative observation. The patterns suggest that, if anything, the control group underreported sensitive behaviors such as stealing. Hence, the treatment effects may actually underestimate therapy's impacts. We also learn that the control group reported fewer expenditures in the survey versus the validation exercise, suggesting that some of the short-term economic gains from cash may be illusory. These insights come with the caveat that they assume that data collected though in depth interviews on a small subset of questions, with a focus on trust-building, are more accurate than survey measures. It is better to think of the validation as a confidence-building exercise rather than hard proof.

In addition to evaluating the pairing of an economic intervention with CBT, this study addresses several gaps in the literature. One is the absence of evidence on behavior change outside the United States, especially in fragile states. Even within the United States, however, there is limited evidence on adult behavior change. Most evaluations of US-based crime and violence-reduction programs focus on education and employment interventions.[5] Studies of CBT tend to be small-sample and nonexperimental (Wilson, Bouffard, and Mackenzie 2005).[6] But STYL's impacts on adult antisocial behavior are consistent with evidence from US adolescents and children, which show that CBT programs in schools and correctional institutes reduce criminal recidivism.[7]

[5] Two US programs, Job Corps and ChalleNGe, are residential programs for at-risk youth that tackle noncognitive skills but focus on jobs and job training (Schochet, Burghardt, and McConnell 2008; Millenky et al. 2012).

[6] An exception is Little, Robinson, and Burnette (1994), who randomly assigned CBT to 1,381 general offenders in Tennessee. They found that re-arrest fell from 56 percent to 41 percent after five years. Our study adds to this large-sample evidence.

[7] For evidence on children and adolescents, see Heckman and Kautz (2014) and Hill et al. (2011). Meta-analyses of adolescent and adult interventions in correctional institutes find that CBT-informed programs appear to outperform alternate therapies (Andrews et al. 1990; Lipsey 2009).

Finally, few studies have measured noncognitive skill and preference changes directly, and so our study strengthens arguments that they are malleable and contribute to antisocial behavior. If we adjust our *p*-values conservatively for multiple comparisons, it is difficult to single out any one skill or preference change, though there is suggestive evidence that time preferences, identity, social networks, and mental health all improve. The malleability of social identity is consistent with evidence from stigmatized Indian sex workers, where short courses of non-CBT psychological therapy increased self-worth, reduced shame, and increased savings and health-seeking behavior (Ghosal et al. 2015). The majority of this evidence, however, comes from small, observational, unpublished studies, which, because of a reliance on administrative data, seldom measure mechanisms directly.[8] But three recent randomized control trials among at-risk Chicago adolescents show that CBT can help adolescents reduce automatic behaviors (such as violent retaliations to a slight) by learning to override *fast* decision-making with conscious *slow* reflection (Heller et al. 2013, 2015). The parallels between that program and STYL, in both the curriculum and impacts, are striking.

It remains to be seen if STYL is replicable, but it is promising that it was adapted from foreign therapies and developed its own facilitators from prior graduates, enhancing scalability. Future work should test the approach in new contexts, compare CBT to other therapies (or a placebo), and reduce the reliance on self-reported data.

I. Intervention and Experiment

Liberia's capital, Monrovia, is home to one-third of the country's 4.3 million people. There are few formal jobs. Most men aged 18 to 35 have limited employment and earn money through a mix of agriculture, casual labor, or petty business. A few turn to crime, which is becoming more violent and commonplace.

During the periods 1989–1996 and 1999–2003, two civil wars wracked Liberia. They killed 10 percent of the population, displaced a majority, and recruited tens of thousands into combat. Since 2003, however, Liberia has been at peace with the help of a United Nations (UN) peacekeeping force. During our study period, 2009–2012, the economy was growing 6 percent per year (Republic of Liberia 2012). Nonetheless, in 2009, people aged 18 to 35 would have spent up to 15 years of their childhood or adolescence in conflict, many robbed of the institutions that normally fostered planfulness, emotional stability, and other noncognitive skills.

Marginalized young men were one of the government's main concerns, especially poorly integrated ex-combatants and other men involved in drugs and crime. Drug and criminal networks are disorganized, but the government worried they could consolidate. They also worried about political violence. High-risk men had joined riots and election violence in the past, and they were targets for mercenary recruitment into the 2010–2011 war in Côte d'Ivoire.

[8] Of 20 studies identified by Wilson, Bouffard, and Mackenzie (2005) only 4 were experimental and 3 of these had sample sizes of 100 or less. The observational studies were also small and were of mixed quality.

A. *Target Population and Recruitment*

We set out to recruit 1,000 high-risk men: men actively involved in crime, interpersonal violence, and drugs, or who were poor and at risk of engaging in these activities. With no administrative data on such men, we recruited them directly. We selected five neighborhoods in Monrovia known for high rates of crime. These were generally mixed-income residential areas with large markets, with populations of roughly 100,000 each.

Recruiters were NEPI affiliates who were not involved in the interventions. NEPI had extensive knowledge of these neighborhoods and connections to local leaders, as well as a strong reputation that target men could verify. Recruiters had worked closely with high-risk men before, and were themselves past graduates of a NEPI program.

We charged the recruiters with finding men who were homeless, drug-using, disreputable in appearance, or present in locations known for crime, armed recruitment, and violence. Community members could easily identify these spots and their denizens. Similarly, certain professions had strong reputations for crime.[9] Appearance was also a useful guide. For instance, recruiters looked for men with a dirty or unkempt appearance, long hair, apparent intoxication, or a "tough" style of dress.

To minimize correlated outcomes and spillovers, we avoided recruiting close associates. We instructed NEPI to approach just one out of every seven to ten high-risk men whom they visually identified. Recruiters then described the therapy, the allocation by lottery, and the baseline survey. They never mentioned cash grants. Over several weeks, recruiters identified roughly 10,000 potentially high-risk men and approached 1,500. Of these, about one-third refused interest in the therapy and survey.[10]

In the end, 999 men agreed to enter the sample. We estimate that they represented 0.6 percent of all adult males in the neighborhoods, and about 12 percent of men aged 18–35 and in the bottom decile of income (online Appendix A.2). Column 1 of Table 1 describes this sample at baseline. On average the men were 25, had nearly 8 years of schooling, earned about $68 in the past month working 46 hours per week (mainly in low skill labor and illicit work), and had $34 informally saved; 38 percent were a former member of an armed group.

B. *Interventions*

Cash.—A nonprofit organization, Global Communities (GC), distributed the cash. They ran a lottery, where winners received $200 cash and losers received a

[9] Location was especially important. Within each of the neighborhoods there were pockets of insecurity where high-risk men were known to live or congregate: abandoned buildings, garbage dumps, drug-dealing spots, parking lots, and homes for rootless young men run by ex-military commanders. Common professions included "car loaders" who have reputations for pickpocketing, or wheelbarrow and motorbike parking areas with reputations for drug selling and crime. They avoided recruiting men known to be "bosses"—men who run homes or drug dens that cater to petty criminals and low-level drug dealers.

[10] We do not have systematic data on refusers, but recruiters reported two main types: men who were poor but were low-risk in that they did not appear to be involved in crime and violence; and high-risk men who said they were too busy to take part in therapy because they had legal or illegal business to attend to.

Table 1—Baseline Summary Statistics and Test of Balance for Select Covariates

	Sample	Test of randomization balance ($N = 999$)						
		Assigned therapy		Assigned cash		Assigned both		F-test
	Mean (1)	Coeff. (2)	p-value (3)	Coeff. (4)	p-value (5)	Coeff. (6)	p-value (7)	p-value (8)
Baseline covariate								
Age	25.40	−0.16	0.68	0.19	0.59	−0.18	0.68	0.18
Married or partnered	0.16	−0.03	0.65	−0.04	0.67	0.04	0.76	0.93
Number of children <15 in household	2.21	−0.58	0.07	−0.51	0.19	0.64	0.29	0.32
Years of schooling	7.72	−0.19	0.68	0.04	0.95	−0.01	0.99	0.55
Has any disabilities	0.08	0.04	0.29	0.00	1.00	−0.04	0.48	0.19
Ex-combatant	0.38	0.06	0.09	0.08	0.08	−0.07	0.11	0.12
Weekly cash earnings (US$)	17.02	−1.89	0.03	−4.85	0.03	5.48	0.00	0.02
Currently sleeping on street	0.24	−0.01	0.80	0.00	0.93	−0.02	0.75	0.26
Savings stock (US$)	33.75	−10.08	0.26	−12.84	0.31	15.69	0.31	0.53
Hours/week, illicit activities	13.55	1.21	0.68	−0.86	0.67	0.06	0.99	0.14
Hrs/week, agriculture	0.36	0.34	0.26	−0.10	0.35	0.13	0.84	0.01
Hrs/week, low-skill wage labor	19.39	0.54	0.88	1.24	0.73	−0.43	0.90	0.94
Hrs/week, in low-skill business	11.53	0.16	0.92	−1.53	0.60	5.76	0.13	0.50
Hrs/week, high-skill work	1.51	−0.05	0.91	0.94	0.03	0.11	0.85	0.01
Sells drugs	0.20	0.01	0.69	0.00	0.92	0.00	0.93	0.92
Uses marijuana daily	0.44	0.08	0.15	0.04	0.13	−0.09	0.22	0.34
Uses hard drugs daily	0.15	−0.04	0.21	0.02	0.52	0.01	0.90	0.37
Committed theft, past 2 weeks	0.53	0.05	0.51	0.01	0.61	−0.02	0.62	0.77
R^2		0.16		0.12		0.35		
p-value on F-statistic		0.50		0.44		0.50		

Notes: We report a selection of covariates here, and all 57 covariates are reported in online Appendix A.1. Column 1 reports the sample mean. A small number of missing values are imputed at the median. Columns 2–7 report the coefficients and p-values from ordinary least squares (OLS) regressions of each baseline covariate on three indicators, one for assignment to each treatment arm, controlling for block fixed effects. Column 8 reports the p-value from a joint test of statistical significance of all three treatment indicators.

consolation prize of $10. There was minimal framing.[11] GC held cash lotteries a week after the end of therapy.

Therapy.—CBT is a short-term approach that tries to reduce self-destructive beliefs or behaviors and promote positive ones. It does so in two ways. First, the therapist tries to help the patient become more aware of their automatic thoughts: inaccurate or negative thinking about themselves or others. Shifting automatic thoughts allows them to respond to everyday situations in a more effective way. A central principle of CBT, however, is that sustained changes in behavior or symptoms also come from actively practicing new behaviors, often starting with simple tasks and, through repetition, positive reinforcement, and gradually increasing the difficulty or complexity of the tasks, changing both behavior and thinking. This practice happens in therapy but also as "homework" (Beck 2011).[12]

[11] See online Appendix B.4 for implementation details. Prior to the lottery, subjects were given about 15 minutes of information on how to keep the money safe (e.g., depositing it with a bank) and examples of what they could use it for (e.g., starting a small business or home improvement). But GC explicitly emphasized to subjects that the grant was unconditional and they were free to do what they wished.

[12] CBT has been studied extensively and validated as a treatment for several of the behaviors targeted by STYL: anger, aggression, criminality, and substance abuse (Saini 2009; Pearson et al. 2002; Wilson, Bouffard, and Mackenzie 2005; Del Vecchio and O'Leary 2004).

Origins and Aims of STYL.—STYL grew out of the experiences of NEPI's founders, but over time they integrated standard Western CBT practices, in part through interactions with international organizations and experts.

The program combined group therapy with one-on-one counseling. Twenty men met in groups three times a week, four hours at a time, led by two NEPI facilitators. On alternate days when groups did not meet, facilitators visited men at home or work to provide advising and encouragement. NEPI offered no compensation except lunch, since men who sacrificed four hours of work could not afford to eat.

NEPI designed the curriculum and approach to encourage two main kinds of change. First, to foster future orientation over present-biased behavior, they taught skills of self-control: to manage anger and emotions, reduce impulsivity, become more conscientious and persevering, and become more planful and goal-oriented.[13] While impulsivity and conscientiousness are often described as personality traits, such traits appear to evolve over the life cycle and are affected by upbringing and investment, so we follow Heckman and Kautz (2014) in considering them skills. These skills of self-control may or may not be related to economic time preferences. The measures and model we outline below treat time preferences and self-control as distinct, and whether they covary is an empirical question.[14]

Second, NEPI tried to persuade men to adopt an anticriminal, antiviolent lifestyle, identity, and values, by helping them self-identify as a normal society member rather than an outcast. A premise of STYL was that the security and respect associated with a mainstream identity were familiar and attractive to the men. So were the values associated with a mainstream identity; it was no mystery that crime and drugs were considered "bad." But to someone who identified as an outcast, those norms and values did not apply to outcasts like them, and a mainstream identity seemed out of reach.

NEPI facilitators tried to persuade the men that this identity was attainable, and that the men should at least try. Partly through exercising skills of self-control, and partly by practice and exposure to new situations, the STYL curriculum walked men through the process of change. The facilitators were an integral part of this intervention, because they modeled the change in skills and values. All were graduates of a prior STYL-like program run by NEPI, and three-quarters were former street youth or combatants.

There are parallels to interventions which show that aspirations—forward-looking goals or targets—influence behavior and respond to intervention (Bernard et al. 2014). There are also parallels to switching social identity.[15]

[13] Note that psychologists also use *self-control* to refer to abilities such as executive function (EF) and delay of gratification (DoG), both of which are thought to lead to less impulsive decision-making and influence long-term success (Mischel, Shoda, and Rodriguez 1989). Some evidence suggests that EF and DoG are distinct from our character skills and are less malleable (Duckworth and Schulze 2011). We measured EF and DoG but they were not the focus of the therapy and we did not hypothesize any change.

[14] The limited evidence comparing economic time preferences to psychological measures of self-control suggests correlations are positive but low, suggesting they may be distinct (Becker et al. 2012).

[15] Akerlof and Kranton (2000) review a wide social science literature. Relatedly, criminologists sometimes refer to a similar process of *knifing off* from old social rules and behaviors, and associate these changes with significant turning points in life, such as marriage, a move, or a life-threatening experience (Maruna and Roy 2007). This literature almost always ties successful knifing off to having a new *script* for the future. The STYL program is effectively that script.

STYL Curriculum and Approach.—The sessions employed a variety of techniques, from lectures and group discussions, to various forms of practice, including: role-playing in class, homework that requires practicing tasks, exposure to real situations, and in-class processing of experiences of executing these tasks. Like many CBT programs, these tasks began simply and got more difficult over time.[16]

In the first three weeks, facilitators encouraged men to try to maintain some new, simple behaviors. This included getting a haircut and removing facial hair, wearing shoes and pants instead of sandals and shorts, improving personal hygiene and the cleanliness of their living area, and reducing substance abuse. These simple exercises in goal-setting and self-control also helped men start to operate within mainstream social norms.

In the middle weeks, facilitators encouraged men to engage with society in planned and unaccustomed ways, akin to exposure therapy.[17] For instance, homework included reintroducing themselves to their family, joining community sports, and visiting banks, supermarkets, shops, and so forth. Men also studied successful people in their community, and reached out to one as a mentor. Men then processed their attempts as a group. Often homework was independent, but facilitators might accompany the more troubled men.

Men also learned to manage emotion: practicing nonaggressive responses to angry confrontations in class, and recognizing signs of angry reactions and learning to distract or calm oneself (walking away, doing other activities, or breathing techniques).

In the last weeks, facilitators taught planning and goal setting. These lessons included training on breaking down large goals into smaller accomplishable subgoals, and then creating plans to accomplish them via concrete steps. For example, men would list subgoals of a plan; these were written on a paper in front of the room, for all to see; the group critiqued them; and plans were rewritten. For homework men would attempt planning in their own lives: how to feed their family the next day; starting a garden; making a savings plan; reconciling with estranged family; or starting a business. These assignments began easy and got more difficult. This process of goal identification and planning is central to most forms of CBT, especially for disruptive behavior disorders (Langberg et al. 2013).

Cost.—The cost of delivering both interventions was $530 per head: $189 for CBT, $216 for the grant, and $125 for registration and administration.

C. *Experimental Design*

We used a 2×2 factorial design. The experiment proceeded as follows. First, following the baseline survey, the respondent was assigned to an offer of therapy

[16] Online Appendix B.3 describes the curriculum in more detail. The full program manual is available at http://chrisblattman.com/documents/policy/2015.STYL.Program.Manual.pdf.

[17] Therapy for patients with social phobia practices similar engagement (Ponniah and Hollon 2008). Besides practice, subjects learn that social feedback is less negative than feared. By reengaging with society, participants tested their negative beliefs about themselves.

by drawing chips from an opaque bag.[18] Therapy began one week later. About one to two weeks after therapy ended, GC announced and held a private draw for $200 grants among the full sample, in blocks of roughly 50 men. Finally, a third organization (Innovations for Poverty Action) ran endline surveys 2 and 5 weeks, and then 12 and 13 months, after grants.

The sample were mistrustful of authority, and we randomized by individual draw rather than computerized assignment to maximize transparency and staff safety. Men in each block took turns drawing colored chips from a fabric bag.[19]

Balance.—This resulted in 25 percent assignment to cash only, 28 percent to therapy only, 25 percent to both, and 22 percent to neither.[20] Treatment is largely balanced along covariates. Table 1 reports tests of balance for each treatment, for selected covariates (see online Appendix A.1 for all, and for endline respondents only). Of 57 covariates over 3 treatments, 14 (8.2 percent) have a difference with $p < 0.05$, and within treatment arms the covariates are not jointly significant. There are, however, large though not statistically significant baseline differences between the control group and the treatment arm that received both therapy plus cash, many in the direction of the *both* arms having better prerandomization outcomes: a random outcome that is in part a result of the smaller sample in these subgroups. To reduce concerns that these imbalances cumulate and could influence endline results, in online Appendix E.8 we create indexes of baseline data much like the endline indexes and look for treatment effects (where these endline behaviors are available) and look for treatment effects. We find that no baseline effect is significant at the 95 percent confidence level, and none of those that are marginally significant ($p < 0.10$) are in the therapy plus cash treatment arm. Therefore, baseline imbalances are unlikely to be driving our results. However, we still control for baseline covariates in our main empirical specification.

Compliance.—Both interventions had high compliance, in part due to NEPI's persuasive efforts and street credibility. Of men assigned to the grant, 98 percent received it. Of men assigned to therapy, 5 percent attended none, another 5 percent dropped out within the first three weeks, and two-thirds attended at least 80 percent of all sessions (online Appendix A.4). Those who dropped out early had less schooling, less self-control, and were less likely to exhibit antisocial behaviors such as substance abuse or stealing (online Appendix A.3). Thus, the highest-risk men seem more likely to have attended over poorer, noncriminal men.

[18] The baseline survey was conducted in central locations, and these draws were systematized and supervised by members of the research team. Only the subject and the research team knew the assigned treatment, however. It was not a public draw.

[19] The order of selection was deliberately unsystematic but not randomized. The number of chips in the bag generally exceeded the number of draws, partly to avoid a correlation between order of the draw and treatment assignment probabilities, and partly to avoid having late-drawing men receive their status by default. See online Appendix B.2 for full details.

[20] See online Appendix A.2 for details by phase and neighborhood. The excess therapy assignments is in part chance, and is in part driven by two blocks where excess treatment chips were accidentally used. All regressions include block fixed effects to account for this.

Phased Implementation.—For logistical reasons we recruited, treated, and studied the men in three phases. A pilot phase recruited 100 men, to ensure that the therapy and cash grant caused no harm, to assess statistical power, and to allow us to refine experimental protocols. The pilot showed no indication of harm, and so we scaled to a further 900 with only minor changes to the interventions and protocols in two phases.[21]

II. Conceptual Framework

The interventions were designed to affect two main outcomes: economic performance and antisocial behaviors (including crime and aggression). To structure this argument, we start with a model of occupational choice between legal and illegal work (such as crime, mercenary work, or election thuggery).[22] Later we consider how such a model could be used to understand other forms of antisocial behavior. We develop the formal model in online Appendix C and outline the structure and results here.

CBT could affect these outcomes in many ways but, as outlined in the next section, we focused on and prespecified three intermediary outcomes: time preferences, self-control, and the values associated with a mainstream social identity. Our simple model examines comparative statics in all three. A change in time preferences is the simplest to examine in a standard model, and we consider the effects of changes in both the discount rate and time inconsistency. More broadly, forms of self-control such as improved emotional regulation, planning, and conscientiousness could be considered a form of human capital that affects productivity. Hence, we also consider what happens when we model self-control as a parameter individual production function. Finally, we introduce a change in criminal identity/values as a change in intrinsic preferences over criminal occupations.[23]

Of course, the therapy is a multifaceted treatment that likely operates through a number of other mechanisms (changed peers or family circumstances, mental health, prosocial preferences) and affects other outcomes and behaviors that themselves are associated with crime (drug abuse or prosocial behavior). We examine these empirically below, but focus the model on the mechanisms that are most in line with NEPI's design principles and theories, as well as the psychological theory and evidence cited above.

Setup.—Suppose people can allocate their time between leisure l, legal work L^b such as petty business or labor, and illegal occupations L^c such as crime, mercenary work, or election thuggery. We refer to these simply as *business* and *crime.*

We assume crime uses labor alone and pays a wage w, which may be uncertain. This resembles the returns we observe to illegal work of the type available to our

[21] See online Appendix A.2 for neighborhood details. Online Appendix B includes the power calculations behind our experimental design.

[22] It is rooted in models of occupational choice with capital infusions and adapted to illicit behavior, as in Blattman and Annan (2016), in the tradition of many economic crime models (Draca and Machin 2015).

[23] Typically, models treat such preferences as fixed, or ignore them. We outline how exogenous changes in noncognitive abilities or preferences affect the comparative statics in an otherwise standard model.

population in Liberia.[24] In the budget function, crime also carries a punishment f with probability ρ, and we assume this risk increases with the time devoted to crime. Punishment could mean prosecution, mob justice, or social sanctions.

Business uses capital, yielding output $F(\theta, L_t^b, K_t)$ where θ is individual ability and K_t is capital inputs. People start with wealth in the form of a riskless asset, a_0, and save or borrow at interest rate r. Self-control skills are one element of θ, and output increases in θ. For simplicity, we focus on the case where self-control skills are inputs into business but not crime. This is the interesting and relevant case, since otherwise investments in self-control skills will not affect occupational choice. We did not assume this from the outset, recognizing that in principle STYL could teach men to be more effective criminals. The pilot phase, however, suggested the opposite was true, so this is the most useful case to discuss.

People choose consumption, labor supply in each sector, and the amount of wealth to invest in business (versus the safe asset) in order to maximize their utility subject to the constraint that consumption plus wealth is equal to total income from business, crime, and the interest on investment. We allow people to be present-biased in the sense that they have a general intertemporal discount factor δ, but can also be time-inconsistent with an extra factor denoted $\beta < 1$ that multiplies all future periods relative to the present (the standard form of quasi-hyperbolic time preferences).

Finally, people value consumption and leisure, but we also allow for a consumption value from conforming to one's identity and values (Akerlof and Kranton 2000; Bénabou and Tirole 2004).[25] In this case, a person's identity and associated values can penalize criminal acts. We use σ to indicate a preference against crime, and we put it in the utility function, $U(c, l, \sigma L^c)$, to distinguish these internal preferences from external punishments f.

We are interested in the effect of the interventions on criminal versus legal labor. Therapy can potentially influence this occupational choice through noncognitive skills θ, time preferences (δ or β), anticriminal values σ, or all of the above. Cash, meanwhile, can influence occupational choice by increasing the assets available for capital inputs into legal business.

Occupational Choice in the Absence of Interventions.—Where financial markets work well and where people are time consistent ($\beta = 1$), businesses are at their optimal scale—they have borrowed until the marginal return to capital is equal to r. Of course, the poor are typically credit-constrained. In this case poor people are forced to invest in capital over time until they reach the same optimal scale. The young and those who have experienced bad shocks will be the furthest behind. As a result, crime is more likely to be chosen by men with low business ability θ, the poor and credit-constrained, those with low disutility of crime, and the time-inconsistent. People may also choose both crime and

[24] Petty crime requires little capital; drug dealers typically work for a "boss" who owns the supply; and those who leave town to work in illicit mining work as "mining boys" for capital-owning "miners" on short-term renewable contracts that pay a daily wage plus a payment tied to output. This is also why we assume below that self-control skills are less important for success in criminal activities.

[25] We ignore the possibility, proposed by Bénabou and Tirole (2004), that ability is imperfectly known and correlated with perceived self-image.

business. Credit-constrained people with partial capital for business may still spend some time in crime. Also, risk averse people may do both activities when returns are uncertain.

Impacts of Cash.—If there are no credit constraints, cash windfalls will not affect occupational choice. But if people are poor and credit-constrained, windfalls will be partly invested in business. People involved in crime will shift to business, especially those with high business ability. Cash infusions will lead to a smaller increase in business for time-inconsistent individuals, however, since they will choose to consume more today.

Impacts of Therapy.—Therapy could increase σ, θ, β, or δ. These channels have distinguishing predictions. Interventions that increase the disutility from crime, σ, will reduce time devoted to it, but will have no effect on returns to business. Interventions that increase noncognitive ability θ will induce more time and investment in business, and also reduce crime. With the presence of risk in both sectors (and assuming risk aversion), interventions in θ will have relatively greater effects in terms of pushing individuals away from crime, because an increase in θ now also makes business relatively less risky. A rise in σ will also have a bigger effect when there is uncertainty, because risk aversion will reinforce the rise in crime aversion and further reduce hours in crime.

What if an intervention increases time consistency, β? This will increase business investment and earnings among the credit-constrained. If people become more time-consistent, they will be more strongly influenced by the consequences of their actions in terms of punishments, and will therefore reduce criminal labor (and increase business labor) as well. Similar comparative statics come from an increase in patience.

Cash and Therapy in Combination.—The model implies that both interventions should lead to a larger decline than one alone simply because the effects are cumulative. Moreover, when people are credit-constrained and also receive cash, this simple model predicts that the effects of a change in σ or θ will be greater with cash than without it. Thus, the interventions may be complementary and the total effect could be greater than the sum of the parts. Note that this simple model does not allow cash to have direct behavioral effects through practice of new behaviors or reinforcement of therapy's lessons.

Relevance of the Model for Aggression.—This model is most useful for thinking about illegal acts that carry material rewards. Other violence does not earn a wage, or does not have an opportunity cost of time. Nonetheless, we can cautiously use the model to think about acts such as aggression. For instance, we can think of some acts as having consumption value that is fleeting (the expressive pleasure of anger or revenge) or persistent (deterring future slights). In this case, $\sigma < 0$. Like crime, these acts carry a risk of punishment. If the criminal wage is zero, there is still a trade-off between the consumption value today and the risk of punishment tomorrow, and the main comparative statics of therapy are similar to the case of crime: instilling values against violence (increasing σ) will reduce aggression; and

increasing time consistency, β, also reduces aggression. Cash, however, will have little deterrent effect on aggression.[26]

III. Data

We tried to survey each subject five times: (i) at baseline prior to the intervention; (ii and iii) at short-run endline surveys 2 and 5 weeks after the grants; and (iv and v) at two endline surveys 12 and 13 months after grants.[27] We ran pairs of surveys to reduce noise in outcomes with potentially low autocorrelation (such as earnings or criminal activity). To measure baseline time preferences and abilities (such as executive function), following each survey the respondents also conducted 45 minutes of incentivized games and tests.[28]

This sample was mobile and difficult to track. Roughly 40 percent changed locations between surveys, many changing sleeping places every few weeks or nights. Just 30 percent had mobile phones. Most went by several aliases, and may have been on the run. To minimize attrition, we collected extensive contact information (all known addresses, plus at least five close contacts), and went to extreme effort to locate each person, wherever they had moved, averaging three to four days of searching per respondent per survey.

We collected data on 92.4 percent across all endline surveys. Attrition is relatively unsystematic: treatment arms had similar response rates (within 0.4 percent of the control group) while a test of joint significance of all baseline covariates yields $p = 0.328$.[29]

We collected longitudinal qualitative data to better understand the context, intervention, and mechanisms. First, a Liberian research assistant acted as a participant-observer during the Phase 1 therapy. Second, we interviewed facilitators for their impressions of the intervention and participants. Third, 3 Liberian research assistants conducted semiscripted interviews, 14 pretreatment and 130 posttreatment, with 66 men in the sample.[30] Interviews covered job satisfaction, investments, economic challenges, plans, antisocial behaviors, and perceptions of the interventions.

[26] In this simple case, there is no role for self-control skills, θ, in aggression. This is a drawback of adapting the pecuniary crime model, since STYL explicitly teaches men skills to regulate their emotions in charged, automatic situations. In some sense, then, STYL may not only change the underlying value of σ (the extent of one's desire not to engage in criminal activity) but also one's ability to ensure that expressed actions conform to the underlying preferences rather than succumbing to immediate temptation or anger. This is functionally equivalent to predictions associated with a larger σ.

[27] The exception is the 100 men in the pilot, which had a single short-run survey 3 weeks after grants. Actual survey times were, on average, 2.2, 5.7, 55.4, and 61.1 weeks after grants. Surveys were 90 minutes long and delivered verbally by enumerators in Liberian English on handheld computers.

[28] See online Appendix D for measurement details. Average winnings equaled about one-half-day's wages.

[29] See online Appendix A.3 for tracking techniques, response rates by survey wave and treatment group, and correlates of attrition. Of the 298 nonresponses (of 3,896), we (i) had no location information (75 percent); men were mentally incapacitated (1 percent), died (8 percent, or 9 men), were in prison (12 percent), or refused (3 percent). Covariates associated with higher attrition include better mental health and income.

[30] Nineteen in control, 16 in therapy, 15 in cash, and 16 in therapy then cash. Sampling was purposeful, based on variation in key baseline measures: economic success, crime, drug use, and present bias.

A. *Key Outcomes and Multiple Comparisons*

After observing the pilot results, we decided to focus on five primary outcomes: two *ultimate* outcomes—antisocial behavior and economic performance—and three intermediary outcomes: economic time preferences, self-control skills, and anticriminal identity/values. The study began before the advent of the social science registry, but we outlined these core hypotheses in a 2012 National Science Foundation (NSF) proposal 1225697.[31] The proposal does not have the level of detail or precision as a pre-analysis plan, but it does describe our main hypotheses and approach to measurement in some detail. The results we present hew closely to the proposal, with only a small number of exceptions.[32]

Naturally, CBT could influence antisocial behavior through other mechanisms, such as drug abuse, changed social networks, or mental health. While plausible, these were not the primary focus of the therapy's design, and as such we did not specify them as core hypotheses in the NSF proposal. These secondary intermediary outcomes are important and relevant, however, and we measured and report on them for this reason.

To organize and reduce the number of hypothesis tests, we combine related measures into mean effects summary indexes.[33] We do so for the two ultimate outcomes of interest plus an index of all intermediary outcomes (primary and secondary). We classify intermediary outcomes into six families: time preferences, self-control skills, identity/values, mental health, substance abuse, and social networks. The first three families were prespecified, while the latter three families (the secondary outcomes) were determined ex post, based on perceived conceptual similarity. Online Appendix D describes these measurement decisions.

Moreover, the tables present unadjusted standard errors as well as p-values adjusted for multiple comparisons. We use the Westfall and Young (1993) free step-down resampling method for the family-wise error rate (FWER), the probability that at least one of the true null hypotheses will be falsely rejected, using randomization inference.[34] Rather than simply adjust for comparisons across the

[31] See Blattman, Jamison, and Sheridan (2012), where the core hypotheses (and division into ultimate and intermediary outcomes) are outlined in Sections 1 and 4.1, and the operationalization (and measurement) of key outcomes in Section 4.4 of the proposal.

[32] These decisions are described in online Appendix D in detail. First, our final measure of antisocial behavior (which we called *crime and violence* in the proposal) does not include political violence, because none occurred before endline. Second, after the proposal but before data collection, we excluded executive function from our measure of self-control, since we determined it was unlikely to be affected by the therapy. Third, after the analysis, we expanded our measure identity and value change to include prosocial behaviors and appearance change, at the suggestion of referees. These changes have only a modest effect on the results, as outlined in online Appendix E.1.

[33] We take averages of our outcome measures, coded to point in the same direction, akin to the approach by Kling, Liebman, and Katz (2007). Note that the outcomes used to create the summary index may themselves be composites of many survey questions, such as consumption (a composite of many goods) or an aggressive behavior index (a composite of many types of aggressive behavior, a standard way that psychologists measure aggression). We do so because it is typically the composite itself rather than its component survey questions that we have theoretical interest or priors. In most cases this is reflected in the survey design, where the survey questions in each composite measure comprise a separate survey section. Also, to create an index by averaging the component variables would give more weight to outcomes that are typically measured with many different questions (such as aggressive behavior) versus one that can be precisely measured with a small number of variables (such as drug selling), which we find inappropriate. Nonetheless, online Appendix E.2 shows robustness to an index that averages all survey questions rather than composite measures, or uses covariance weighting rather than mean effects.

[34] Other papers taking this approach include Kling, Liebman, and Katz (2007); Casey, Glennerster, and Miguel (2012); and Anderson (2008). Using the Westfall-Young bootstrap and the Holm-adjusted Bonferroni methods yields similar results.

major family outcomes, we also adjust for the fact that we are estimating three treatment effects (one for each arm). Thus, for our three main family indexes (economic behavior, antisocial behavior, and intermediary mechanisms) we report *p*-values adjusted for nine comparisons in total. By reporting both the adjusted and unadjusted statistical significance, readers can use the threshold appropriate to their question and preferences. If interested in the specific hypothesis of the effect of CBT on antisocial behavior, the unadjusted *p*-value is appropriate. If asking about the effects of different treatment combinations on mechanisms, as in our NSF proposal, the conservative adjusted *p*-values are more appropriate.

IV. Empirical Strategy and Estimation

We estimate intent-to-treat (ITT) effect on outcomes, Y, via the OLS regression

$$(1)\quad Y_{ij} = \tau_1 TherapyOnly_i + \tau_2 CashOnly_i + \tau_3 Cash\&Therapy_i + \mathbf{X}_i\lambda + \gamma_j + \varepsilon_{ij},$$

where *TherapyOnly*, *CashOnly*, and *Cash&Therapy* are indicators for random assignment to treatment arms: therapy only, cash only, or both therapy and cash. We control for a vector of baseline characteristics, $\mathbf{X}$, and fixed effects for each of the j randomization blocks, γ_j; Y_{ij} is the average of the two proximate survey rounds (e.g., the two- and five-week surveys for short-term effects). To reduce sensitivity to outliers, we top-code continuous variables at the ninety-ninth percentile. We test sensitivity to alternative approaches in online Appendix E.2.

Self-Reported Data.—One threat to identification comes from systematic measurement error in self-reported data, especially measurement error correlated with treatment status. In the absence of administrative data such as arrest records, we developed a technique to validate select survey variables through intensive observation. Blattman et al. (2016) reports the approach in detail, and we summarize in Section VI and online Appendix F.

Spillovers.—Another threat to identification comes from spillovers. Our recruiting strategies—working in large neighborhoods, recruiting less than 1 percent of adult men in those areas, and less than 15 percent of high-risk men we could identify on the street—were designed to reduce equilibrium effects such as a change in the returns to illicit work. We do not have the data or research design, however, to confirm that these equilibrium effects were minimized.

Another potential spillover involves interactions within and between treatment arms, especially therapy. For example, because of peer effects and the emphasis on social norms, there could be positive spillovers from treating groups of friends. If so, the coefficients on therapy would overestimate the effect of therapy in isolation. Alternatively, to the extent that control subjects interact with and learn from treatment subjects, they may acquire some of the lessons, leading us to underestimate therapy's impact.

We designed recruitment to minimize such interaction bias, but could not eliminate it. We do not have detailed social network data for the full sample, but we did trace social networks within the first two therapy groups. On average, each subject

was acquainted with 6 of the 43 others in therapy, and 30 percent reported 1 close associate in therapy. Given randomization, we can assume similar relationships in the other arms. Without systematic data on networks we cannot estimate spillovers, and this is a weakness of our design. The two effects should cancel each other out somewhat, but the extent is unknown.

Interpretation and Generalizability—Another point is that our sample is not drawn from a well-defined population. This is a function of the setting—there is no administrative record of high-risk men in Liberia (or in any low-income or fragile state). We recruited men in a relatively transparent, replicable fashion, but one-third declined to enter the study for reasons we cannot observe. Thus, the treatment effects we estimate cannot be generalized to a defined population. This is not only a constraint of the setting, but also the nature of a proof-of-concept trial, where we have two promising but highly uncertain treatments—unconditional cash and CBT. Thus, our study is akin to a medical efficacy trial, to determine whether the intervention produces the expected result under favorable circumstances.

V. Results

Figure 1 reports ITT estimates using equation (1) on the two ultimate outcomes of interest and an index of all intermediary outcomes. Figure 2 reports ITT estimates for the six intermediary outcome families. Both panels display regular 95 percent confidence intervals as well as *p*-values unadjusted and adjusted for multiple comparisons, as outlined in Section IIIA. Figure 1 adjusts for 9 comparisons (3 arms $\times$ 3 outcomes) and Figure 2 adjusts for 18.[35]

Broadly speaking, cash did not lead to a statistically significant or sustained reduction in overall antisocial behaviors, but therapy did. In the short run, therapy led to large reductions, by 0.25 standard deviations with therapy alone and 0.31 standard deviations with therapy plus cash. This reduction in antisocial behaviors persisted, however, only when therapy was followed by cash: one year after therapy, therapy alone led to a 0.08 standard deviation fall in antisocial behaviors (not statistically significant) compared to a 0.25 standard deviation fall with therapy plus cash (significant at the 1 percent with unadjusted *p*-values, and at the 5 percent level if adjusted). This difference between therapy and therapy plus cash after 12– 13 months is significant at the 5 percent level using unadjusted *p*-values.[36]

[35] These are calculated separately for the 2–5 week and 12–13 month surveys. Online Appendix E.1 also displays results if we ignore the ultimate/intermediary and primary/secondary distinctions, and simply adjust *p*-values for $3 \times 8 = 24$ comparisons. In this case, the short-term conclusions are unaffected, but over 12–13 months, the effect of cash plus therapy on antisocial behaviors has a *p*-value of 0.106. As seen in online Appendix E.2, these effects are robust to a variety of specifications and attrition scenarios. We obtain similar results if we: pool the endlines rather than averaging them; construct summary indexes of all underlying survey questions rather than indexes of the composite measures; or covariance weight rather than weight index components equally. We also show that the results are robust to conservative attrition scenarios by substituting extreme values for missing outcomes.

[36] See online Appendix E.3 for formal tests of the difference between both therapy and cash to therapy or cash alone. Online Appendix E.4 tests whether 2–5-week and 12–13-month impacts are equal. We cannot reject the hypothesis that the effects of both therapy and cash are equal over time, but we can reject the equality of effects for therapy alone.

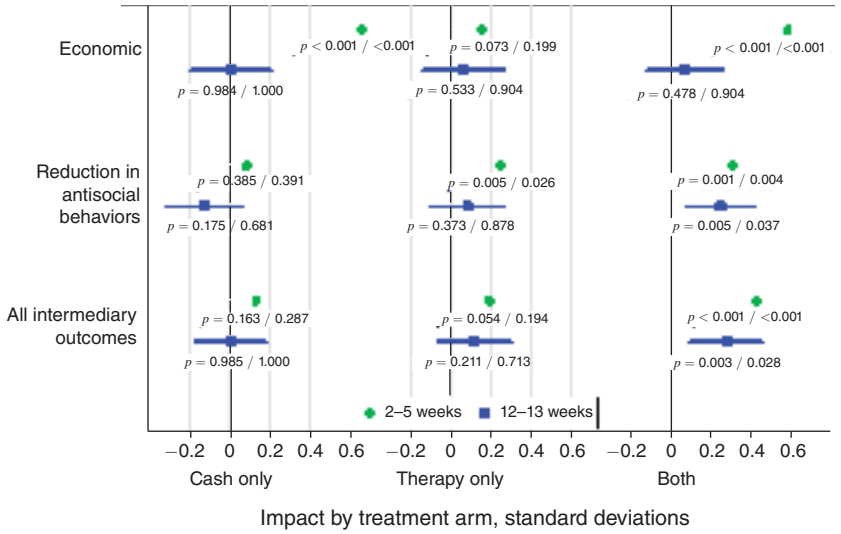


FIGURE 1. PROGRAM IMPACTS ON THE TWO ULTIMATE OUTCOMES AND AN INDEX OF ALL INTERMEDIARY OUTCOMES (z-SCORES) WITH 95 PERCENT CONFIDENCE INTERVALS AND UNADJUSTED/ADJUSTED p-VALUES

Notes: The figure reports the effect of each treatment arm after 2–5 weeks and 12–13 months with 95 percent confidence intervals and two p-values, one unadjusted and one adjusted for 9 comparisons (3 arms and 3 outcomes) using the Westfall-Young method. Treatment effects are estimated via OLS controlling for baseline covariates and block fixed effects. Each summary index is the standardized mean of composite outcomes. Standard errors are heteroskedastic-robust.

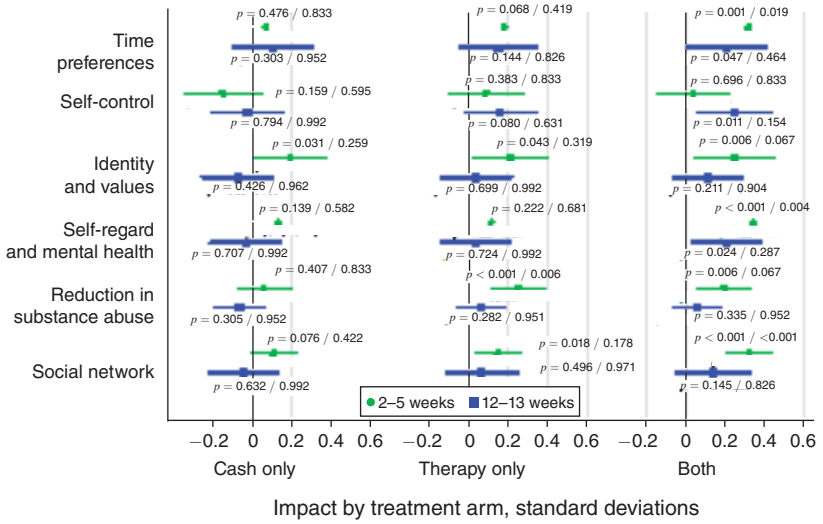


FIGURE 2. PROGRAM IMPACTS ON FAMILIES OF INTERMEDIARY OUTCOMES (z-SCORES) WITH 95 PERCENT CONFIDENCE INTERVALS AND UNADJUSTED/ADJUSTED p-VALUES

Note: Estimated as in Figure 1, except that adjusted p-values account for 18 comparisons (3 arms and 6 outcomes).

We see a change of similar proportions of the intermediary outcomes in aggregate, and this too is only persistent in the group that received therapy plus cash.[37] Individually, most of the intermediary outcomes shift in the expected direction, and moderate over time. In the short run, only time preferences, mental health, and social networks are statistically significant using the more conservative adjusted *p*-values. After a year, no individual intermediary outcome is significant using the adjusted *p*-values, although all (except substance abuse) are pointing in the expected direction.

A. *Antisocial Behaviors*

We defined antisocial behaviors as disruptive or harmful acts toward others, such as crime or aggression. Thus, the family excludes self-harm (e.g., drug abuse) or acts by peers. Table 2 reports impacts on the components of the antisocial behavior index, for illustrative purposes. The table reports both unadjusted and adjusted *p*-values. The adjusted *p*-values on the antisocial behaviors family index come directly from Figure 1 (adjusted for the nine comparisons of three arms and three outcomes). The adjusted *p*-values on the seven components reflects the $3 \times 7 = 21$ comparisons across all arms and components.

We constructed the seven component measures from sets of related survey questions, each typically from its own survey module. All are self-reported. In general, the coefficients on therapy only or therapy plus cash are negative, and a majority are statistically significant using unadjusted standard errors. We must interpret the individual point estimates with caution, since almost none of the individual components are significant when we adjust for 24 comparisons, save aggressive behaviors.

- *Drug selling and other crime.* In the short run, 17 percent of the control group said they sold drugs, and they admitted to 2.6 acts of theft in the past two weeks. A year later, 13.5 percent sold drugs and they reported 1.9 acts of theft. Crime rates may fall because we are recruiting people in hard times, and there is regression to the mean. With therapy, however, crime rates fell by almost 50 percent in the short run, and this fall persisted a year with therapy plus cash. Online Appendix D describes specific crime measures. To give a crude sense of magnitude, if we extrapolate results to the full year since baseline, therapy plus cash led men to go from 66 to 40 crimes per person per year (online Appendix E.6). Given the $530 intervention cost, this is roughly $21 per crime in the first year, ignoring any ongoing impact on crime or other program benefits.
- *Fights.* We also asked about nine types of verbal and physical altercations in the past two weeks, including the frequency and severity of disputes with peers, neighbors, leaders, or police. Here, as with all summary indexes in the paper,

[37] A natural question is whether the therapy is impactful for the most or least antisocial men. Online Appendix E.7 reports ITT regressions where we add an interaction between treatment arms and a standardized index for baseline antisocial behavior, as well as initial future orientation and self-control. The therapy was impactful for the average participant, but the greatest decline in antisocial behavior was among those with the highest antisocial behaviors and the lowest levels of self-control and future orientation. These estimates must be taken with caution, since the heterogeneity analysis was not prespecified. But these were the only heterogeneity analyses run on antisocial behavior.

Table 2—Program Impacts on Antisocial Behaviors

		ITT regression: ($N = 947$)											
		Therapy only				Cash only				Both			
				p-value				*p*-value				*p*-value	
Outcome	Control mean (1)	ITT (2)	SE (3)	Unadj. (4)	Adj. (5)	ITT (6)	SE (7)	Unadj. (8)	Adj. (9)	ITT (10)	SE (11)	Unadj. (12)	Adj. (13)
Panel A. 2–5 week impacts													
Antisocial behaviors, *z*-score	0.151	−0.249	[0.088]	**0.005**	**0.026**	−0.079	[0.091]	0.385	0.391	−0.308	[0.089]	**0.001**	**0.004**
Usually sells drugs	0.166	−0.077	[0.027]	**0.005**	0.082	−0.041	[0.029]	0.157	0.803	−0.076	[0.028]	**0.006**	0.099
# of thefts/robberies in past 2 weeks	2.577	−0.841	[0.400]	**0.036**	0.359	−0.770	[0.409]	0.060	0.502	−1.236	[0.407]	**0.002**	**0.045**
Disputes and fights in past 2 weeks, *z*-score	0.076	0.013	[0.092]	0.888	0.999	0.027	[0.091]	0.768	0.999	−0.132	[0.076]	0.084	0.596
Carries a weapon on body[a]	0.157	−0.086	[0.034]	**0.013**	0.167	−0.045	[0.037]	0.224	0.888	−0.093	[0.035]	**0.007**	0.115
Arrested in past 2 weeks	0.139	−0.011	[0.027]	0.674	0.999	0.006	[0.027]	0.819	0.999	−0.013	[0.029]	0.637	0.999
Aggressive behaviors, *z*-score	0.102	−0.208	[0.081]	**0.011**	0.151	0.008	[0.085]	0.928	0.999	−0.196	[0.087]	**0.024**	0.275
Verbal/physical abuse of partner, *z*-score[a]	−0.035	−0.087	[0.111]	0.430	0.985	0.091	[0.114]	0.422	0.985	−0.032	[0.115]	0.777	0.999
Panel B. 12–13 month impacts													
Antisocial behaviors, *z*-score	0.032	−0.083	[0.093]	0.373	0.878	0.132	[0.097]	0.175	0.681	−0.247	[0.088]	**0.005**	**0.037**
Usually sells drugs	0.135	−0.034	[0.029]	0.249	0.909	0.035	[0.030]	0.244	0.909	−0.059	[0.029]	**0.041**	0.474
# of thefts/robberies in past 2 weeks	1.839	0.073	[0.395]	0.855	0.997	0.352	[0.388]	0.365	0.943	−0.728	[0.363]	**0.045**	0.482
Disputes and fights in past 2 weeks, *z*-score	−0.060	−0.026	[0.091]	0.772	0.997	0.100	[0.090]	0.267	0.909	−0.100	[0.077]	0.194	0.881
Carries a weapon on body[a]	0.148	−0.059	[0.031]	0.061	0.553	0.043	[0.035]	0.215	0.894	−0.066	[0.033]	**0.049**	0.490
Arrested in past 2 weeks	0.118	−0.006	[0.024]	0.811	0.997	0.007	[0.025]	0.793	0.997	−0.033	[0.024]	0.171	0.863
Aggressive behaviors, *z*-score	0.188	−0.153	[0.110]	0.163	0.863	−0.043	[0.107]	0.685	0.997	−0.339	[0.109]	**0.002**	**0.035**
Verbal/physical abuse of partner, *z*-score[a]	−0.071	0.142	[0.100]	0.156	0.863	0.233	[0.113]	**0.040**	0.474	0.059	[0.104]	0.574	0.992

Notes: The table reports intent-to-treat estimates of the effect of each treatment arm after 2–5 weeks and 12–13 months, controlling for baseline covariates and block fixed effects. We focus on predefined composite measures, typically defined by survey module. For instance, thefts/robberies is the sum of 8 kinds of crimes; disputes/fights is the standardized mean of 9 kinds of physical or verbal altercations with peers, community, and authorities; aggressive behaviors is the standardized mean of 19 possible types of aggression and hostility; and verbal and physical abuse of partners is the standardized mean of 3 forms of verbal abuse of intimate partners plus one form of physical abuse. (For the latter two cases, we report standardized indexes since the incidents are measured on a 0–3 frequency scale, and the absolute sum itself has no interpretation.) The overall summary index is the standardized mean of these seven composite outcomes, standardized. Heteroskedastic robust standard errors are reported in brackets. Adjusted *p*-values use the Westfall-Young method to correct for multiple comparisons, as described in Section IIIA. The overall family index is adjusted for nine comparisons, as in Figure 1. Within each endline round, the component indexes are adjusted for 21 comparisons, for 3 arms and 7 outcomes. *p*-values less than 0.05 are bolded.

[a]These variables were not collected during every phase/round, so their regressions have a smaller sample size.

we use the standardized mean effects of all nine survey questions.[38] On average, men reported about one dispute in the past two weeks. None of the effects are distinguishable from zero, and only the point estimates on therapy and cash are negative.

[38] A main reason is because the measurement scales differ across component survey variables and the absolute valuer of the scales themselves are not meaningful (e.g., a frequency scale of 0–3, from never to often). We standardize individual survey questions, average them, and standardize this composite to have mean zero and unit standard deviation. Results are robust to alternate weighting and indexing approaches.

- *Weapons.* We asked men if they carried a weapon on their body for protection. This was typically a knife, as guns were rare. After a year, 15 percent were carrying a weapon, and this fell about by about one-half with either therapy alone or therapy plus cash.
- *Arrests.* Fourteen percent of the control group reported an arrest in the two weeks before the short-run endline, and 12 percent after a year. We did not see a statistically significant decline in arrests, though after one year the coefficient on therapy plus cash represented decline of almost one-third, or about one arrest per year.
- *Aggressive and hostile behaviors.* We asked 19 questions about reactive and proactive aggression, such as the frequency with which they yell, curse, bully others, cheat, or lose their tempers.[39] After a year, the index of all 19 questions fell 0.15 standard deviations (not significant) with therapy alone and 0.34 with both (significant at the 5 percent level with multiple comparison adjustment).
- *Intimate partner abuse.* We have a crude measure of intimate partner abuse—three questions on verbal abuse (e.g., cursing and yelling) and one on physical abuse in the past two weeks. A standardized index of these measures fell little in the short run with therapy, and after a year the coefficient are actually positive (the only instance where therapy is positively correlated with violence).
- *Political violence.* Given Monrovia's history of mercenary recruitment, riots, and election violence, we predicted the men would have opportunities for political violence. Indeed, shortly after the Phase 1 men received therapy, there was a minor riot in the city.[40] From then, however, Liberia entered one of the most politically quiescent periods in recent history, and so we had no political violence to measure. This is the only prespecified outcome that we could not test directly.

B. *Economic Performance*

Table 3 reports program impacts on an index of measures of economic performance: income, homelessness, savings, investment, and employment levels. In the month after grants, general economic activity increased among those receiving cash alone (0.66 standard deviations) or cash following therapy (0.58 standard deviations). But after a year the effects in all three arms have approached zero. The same patterns hold if we look at income alone.[41] We measured income in three ways: (i) consumption in the past two weeks; (ii) estimated earnings in all activities in the past two weeks; and (iii) an index of durable assets.[42] The short-term rise in

[39] We used nine questions from a standard scale, adapted to Liberian English (Raine et al. 2006), and added ten more locally relevant acts based on our qualitative interviews.

[40] The men in all three treatment arms were slightly less likely to participate or sympathize with the rioters, but with a sample size of just 100 these effects were not significant.

[41] We focus on an overall index of economic performance to reduce concerns of cherry-picking, and because the NSF proposal we use as a *pre-analysis plan* did not single out income as our primary metric.

[42] To obtain earnings, we first asked each respondent their gross and net earnings in the past four weeks across 25 economic activities (legal and illegal). This earnings measure could still be subject to recall and other biases, and may inadequately capture home production. Thus, we also use two measure of permanent income. One is an index of durable assets—a z-score constructed by taking the first principal component of 42 measures of land, housing quality, and small and large household assets. We also conduct an abbreviated consumption module of short-term food and nonfood consumption.

TABLE 3—PROGRAM IMPACTS ON ECONOMIC PERFORMANCE

		ITT regression: ($N = 947$)											
		Therapy only				Cash only				Both			
				p-values				*p*-values				*p*-values	
Outcome	Control mean (1)	ITT (2)	SE (3)	Unadj. (4)	Adj. (5)	ITT (6)	SE (7)	Unadj. (8)	Adj. (9)	ITT (10)	SE (11)	Unadj. (12)	Adj. (13)
Panel A. 2–5 week impacts													
Economic performance, *z*-score	−0.312	0.155	[0.086]	0.073	0.199	0.658	[0.096]	**<0.001**	**<0.001**	0.580	[0.096]	**<0.001**	**<0.001**
Consumption, past two weeks, US$	44.394	10.683	[3.812]	**0.005**	0.066	24.040	[4.199]	**<0.001**	**<0.001**	20.695	[3.553]	**<0.001**	**<0.001**
Weekly earnings in past two weeks, US$	14.410	1.351	[1.592]	0.396	0.907	4.128	[1.572]	**0.009**	0.106	3.036	[1.535]	**0.048**	0.383
Durable consumption assets, *z*-score	−0.093	0.124	[0.094]	0.185	0.738	0.138	[0.094]	0.141	0.672	0.240	[0.097]	**0.014**	0.141
Savings stock (US$)[a]	45.957	−1.652	[8.901]	0.853	0.981	16.342	[10.018]	0.103	0.591	18.131	[10.169]	0.075	0.503
Investment in past two weeks (US$)	18.444	6.748	[6.291]	0.284	0.849	55.750	[8.410]	**<0.001**	**<0.001**	48.122	[7.797]	**<0.001**	**<0.001**
Hours/week of work in past month	36.773	1.044	[2.879]	0.717	0.978	7.439	[2.907]	**0.011**	0.119	1.787	[2.966]	0.547	0.953
Homeless in past two weeks	0.202	0.004	[0.029]	0.882	0.981	−0.090	[0.029]	**0.002**	**0.032**	−0.100	[0.029]	**0.001**	**0.010**
Home robbed or belongings stolen in past month	0.859	−0.011	[0.035]	0.759		−0.009	[0.036]	0.808		−0.012	[0.038]	0.752	

(*Continued*)

consumption is significant at the 1 percent level, and the rise in assets at the 10 percent level, after adjusting for multiple comparisons. An overall index of all three income measures is significant at the 1 percent level (not shown). Homelessness also falls significantly as income rises, but there is no decline after a year.

Consumption and assets could rise simply from spending the grant. But this does not explain the temporary earnings boost. Overall, the cash seems to have been invested in petty business, and this accounts for the rise in short-run earnings. But bad shocks, especially theft, meant these gains were fleeting.

To see this, we assessed grant spending in two ways. Using pictures of different types of spending and plastic chips, we first asked grant recipients to indicate how they used the grant. Table 4 reports self-reported allocations of the grant by treatment arm. We see little effect of the recent therapy on allocations. Little of the grant seems to have been spent on drugs, alcohol, gambling, and prostitution. Even if men underreport these expenses, we see no difference between cash recipients who did and did not receive therapy.

We can also look at expenditure data, which included a range of business investments in the two weeks prior to the two- and five-week surveys. As reported in Table 3, those who received only cash reported $56 more investment in each two-week period. Thus, the total 5-week investment treatment effect is at least $112—just

TABLE 3—PROGRAM IMPACTS ON ECONOMIC PERFORMANCE (*Continued*)

		ITT regression: ($N = 947$)											
		Therapy only				Cash only				Both			
				p-values				*p*-values				*p*-values	
Outcome	Control mean (1)	ITT (2)	SE (3)	Unadj. (4)	Adj. (5)	ITT (6)	SE (7)	Unadj. (8)	Adj. (9)	ITT (10)	SE (11)	Unadj. (12)	Adj. (13)
Panel B. 12–13 month impacts													
Economic performance, *z*-score	−0.031	0.065	[0.104]	0.533	0.904	0.002	[0.101]	0.984	1.000	0.069	[0.098]	0.478	0.904
Consumption, past two weeks, US$	47.347	−1.870	[3.808]	0.623	0.999	−2.615	[3.605]	0.468	0.998	−5.005	[3.589]	0.164	0.934
Weekly earnings in past two weeks, US$	17.595	0.764	[1.723]	0.658	0.999	1.341	[1.663]	0.420	0.998	0.436	[1.882]	0.817	0.999
Durable consumption assets, *z*-score[a]	0.009	0.143	[0.096]	0.138	0.904	−0.111	[0.097]	0.251	0.983	0.086	[0.096]	0.369	0.996
Savings stock (US$)[a]	51.395	11.360	[11.042]	0.304	0.991	2.026	[10.245]	0.843	0.999	21.744	[11.352]	0.056	0.616
Value of business assets (US$)[a]	26.121	3.116	[13.431]	0.817	0.999	19.391	[15.276]	0.205	0.964	13.548	[13.007]	0.298	0.991
Hours/week of work in past month	34.273	1.030	[2.550]	0.686	0.999	0.681	[2.525]	0.788	0.999	−1.416	[2.493]	0.570	0.999
Homeless in past two weeks	0.145	0.018	[0.030]	0.554	0.999	0.016	[0.029]	0.586	0.999	−0.020	[0.030]	0.497	0.999
Home robbed or belongings stolen in past month	0.797	−0.037	[0.043]	0.386		−0.086	[0.042]	**0.039**		0.013	[0.042]	0.766	

Notes: The table reports intent to treat estimates of the effect of each treatment arm controlling for baseline covariates and block fixed effects. The income summary index is the standardized mean of three composite outcomes (themselves first standardized). Heteroskedastic robust standard errors are reported in brackets. Adjusted *p*-values use the Westfall-Young method to correct for multiple comparisons, as described in Section IIIA. The overall family index is adjusted for nine comparisons, as in Figure 1. Within each endline round, the component indexes are adjusted for 21 comparisons, for 3 arms, and 7 outcomes; *p*-values less than 0.05 are bolded.

[a]These variables were not collected during every phase/round, so their regressions have a smaller sample size.

TABLE 4—SELF-REPORTED ALLOCATION OF THE GRANT, BY EXPENDITURE CATEGORY

	Treatment arm		Difference ($N = 475$)	
Expenditure category	Cash and therapy (1)	Cash only (2)	Coeff. (3)	*p*-value (4)
Living expenses (such as food, clothing, rent)	28%	26%	0.03	0.23
Durable assets	7%	6%	−0.01	0.48
Drugs, alcohol, gambling, and sex	4%	4%	0.00	1.00
Gifts and transfers to others	11%	11%	0.00	0.97
Business investments and expenses	23%	24%	−0.03	0.28
Savings and debt payments	20%	20%	0.01	0.56
Own health and education	8%	8%	0.01	0.68

Notes: Using pictures of different types of spending and plastic chips, grant recipients indicated how they used the grant. Columns 1–2 report the means for each treatment arm. Columns 3–4 report the coefficients and *p*-values from an OLS regression of the proportion spent on an indicator for assignment to therapy then cash controlling for block fixed effects and baseline covariates.

over one-half of the grant (significant at the 1 percent level using adjusted *p*-values). Meanwhile, the therapy only group resembled the control group in terms of investment.

These short-run investments did not last. In the cash only group, the stock of business assets after a year was only $19 greater than in the control group, not statistically significant. We also see no one-year difference in total work hours.[43]

What happened? From qualitative interviews, insecure property rights were a major barrier to capital accumulation. A large number of men reported the theft of all their assets, or all their wares, on a regular basis, by criminals or (for market wares) the police.[44] We added this question to the survey (though not as part of the performance index). At each survey round, about 70 percent of the men reported a house robbery and belongings stolen in the past month.[45] This implies a robbery every other month, at least. There is little difference by treatment status, suggesting that men were not more likely to be targeted if they received cash. But they would have had more to lose.

C. *Indications of Noncognitive Skill and Preference Change*

Tables 5 and 6 report treatment effects on our six intermediary outcome families and their components. For the six family indexes, we report regular standard errors as well as *p*-values adjusted for 18 comparisons (3 arms and 6 families), as in Figure 1 above. (Correcting for just the three primary outcomes yields qualitatively similar conclusions.) We also report and discuss the components of each index mainly for illustrative purposes. We add *p*-values adjusted for multiple comparisons within each family, so that a family with 4 components is adjusted for 12 comparisons.

Time Preferences.—We report a summary index of four measures of patience and four of time inconsistency—akin to δ and β in our model. Specifically, we measured: a set of incentivized trade-offs between modest amounts of money now versus in two weeks, and again in two versus four weeks, that allow us to place men in seven ordered bins of patience and time-inconsistency (for an average payout of $3, about a day's wages); a hypothetical (non-incentivized) version of the same trade-offs, with higher stakes trade-offs; and self-reported assessments of time preferences. All are described in online Appendix D.3.

In the short term, time preferences become more forward-looking for all treatment arms, though the result is largest (0.32 standard deviations) and statistically significant only for therapy plus cash (at the 1 percent level even accounting for multiple comparisons). After a year, the point estimates from therapy are positive—0.15 standard deviations for therapy alone and 0.21 standard deviations for therapy plus cash. The latter is statistically significant using regular standard errors

[43] All three treatments caused individuals to substitute from illicit work to nonagricultural low-skill business in the short term, but the effects were most pronounced and longer lasting for the group that received both cash and therapy (see online Appendix E.8).

[44] In some cases this was theft by a friend, family member, or stranger. Also common was confiscation of wares by the police. Some forms of market selling contravene official rules, often unenforced, but nonetheless giving police opportunities to confiscate. Some confiscation is legitimate, some not.

[45] We do not include this outcome in the economic performance index as it's not a measure of economic performance. Rather we report it in the table mainly for descriptive purposes.

TABLE 5—PROGRAM IMPACTS ON TIME PREFERENCES, SELF CONTROL SKILLS, AND IDENTITY/VALUES

		ITT regression: ($N = 947$)											
		Therapy only				Cash only				Both			
				p-value				*p*-value				*p*-value	
Outcome (*z*-score)	Control mean (1)	ITT (2)	SE (3)	Unadj. (4)	Adj. (5)	ITT (6)	SE (7)	Unadj. (8)	Adj. (9)	ITT (10)	SE (11)	Unadj. (12)	Adj. (13)
Panel A. 2–5 weeks													
Forward-looking time preferences	−0.202	0.179	[0.098]	0.068	0.419	0.071	[0.099]	0.476	0.833	0.318	[0.099]	**0.001**	**0.019**
Patience	−0.093	0.187	[0.073]	**0.010**	**0.046**	0.116	[0.073]	0.114	0.280	0.267	[0.074]	**<0.001**	**0.002**
Time inconsistency	0.008	−0.063	[0.074]	0.393	0.596	−0.009	[0.076]	0.903	0.902	−0.138	[0.075]	0.067	0.209
Self-control skills	−0.037	0.085	[0.098]	0.383	0.833	−0.147	[0.104]	0.159	0.595	0.037	[0.096]	0.696	0.833
Impulsiveness	−0.010	−0.011	[0.101]	0.915	0.995	0.180	[0.108]	0.094	0.619	0.104	[0.095]	0.275	0.919
Conscientiousness	−0.077	0.109	[0.105]	0.301	0.919	0.046	[0.106]	0.664	0.995	0.163	[0.105]	0.120	0.684
Perseverance/GRIT	−0.035	0.027	[0.099]	0.785	0.995	−0.130	[0.105]	0.217	0.873	0.042	[0.104]	0.686	0.995
Reward responsiveness	−0.010	−0.071	[0.106]	0.503	0.980	0.107	[0.107]	0.321	0.919	0.013	[0.105]	0.901	0.995
Identity and values	−0.169	0.192	[0.095]	**0.043**	0.319	0.199	[0.092]	**0.031**	0.259	0.268	[0.097]	**0.006**	0.067
Attitudes toward use of violence	0.100	−0.206	[0.094]	**0.028**	0.123	−0.187	[0.096]	**0.051**	0.176	−0.180	[0.097]	0.065	0.176
Index of appearance	−0.118	0.085	[0.081]	0.295	0.298	0.131	[0.081]	0.105	0.189	0.203	[0.080]	**0.011**	0.062
Panel B. 12–13 months													
Forward-looking time preferences	−0.149	0.149	[0.102]	0.144	0.826	0.105	[0.102]	0.303	0.952	0.209	[0.105]	**0.047**	0.464
Patience	−0.240	0.170	[0.103]	0.097	0.350	0.145	[0.096]	0.132	0.386	0.258	[0.099]	**0.009**	**0.048**
Time inconsistency	0.129	−0.072	[0.083]	0.386	0.712	0.018	[0.087]	0.833	0.836	−0.059	[0.084]	0.480	0.712
Self-control skills	−0.070	0.159	[0.090]	0.080	0.631	−0.025	[0.095]	0.794	0.992	0.244	[0.095]	**0.011**	0.154
Impulsiveness	0.082	−0.178	[0.096]	0.064	0.431	0.006	[0.098]	0.951	0.961	−0.212	[0.099]	**0.032**	0.265
Conscientiousness	0.018	−0.065	[0.097]	0.506	0.961	−0.028	[0.100]	0.779	0.961	0.044	[0.097]	0.648	0.961
Perseverance/GRIT	−0.037	0.116	[0.099]	0.241	0.851	0.057	[0.099]	0.565	0.961	0.105	[0.103]	0.311	0.903
Reward responsiveness	0.072	−0.165	[0.102]	0.106	0.580	0.084	[0.100]	0.397	0.938	−0.242	[0.102]	**0.018**	0.177
Identity and values	−0.021	0.013	[0.089]	0.882	0.992	−0.101	[0.089]	0.255	0.962	0.034	[0.090]	0.704	0.904
Anticriminal/ antiviolent values	0.070	−0.076	[0.088]	0.386	0.880	0.026	[0.088]	0.768	0.948	−0.177	[0.086]	**0.040**	0.279
Index of appearance	0.016	−0.102	[0.078]	0.191	0.750	−0.085	[0.077]	0.269	0.813	−0.109	[0.082]	0.182	0.750
Prosocial behavior	0.018	0.041	[0.088]	0.636	0.948	−0.075	[0.085]	0.378	0.880	−0.017	[0.090]	0.850	0.948

Notes: The table reports intent to treat estimates of the effect of each treatment arm controlling for baseline covariates and block fixed effects. We focus on predefined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heteroskedastic robust standard errors are reported in brackets. Adjusted *p*-values use the Westfall-Young method to correct for multiple comparisons, as described in Section IIIA. The overall family index is adjusted for 18 comparisons, as in Figure 2. Within each endline round, the component indexes are adjusted for three arms and the number of components within the family; *p*-values less than 0.05 are bolded.

but, like all the family indexes, is not significant after a year after accounting for multiple comparisons. Looking within the family index, point estimates are larger and more precise for patience than time inconsistency.

Self-Control Skills.—We measured self-control skills using standard psychometric questionnaires for four constructs that psychologists associate with less impulsive and more planful behavior.[46] First, we looked at nine questions from the Barrett Impulsiveness Scale (Spinella 2007), which assesses one's inability to

[46] In addition to these four psychological scales, we also conducted tests of executive function—cognitive processes associated with inhibitory control, working memory, self-regulation, and planning, such as digit recall (see online Appendix D.6). We did not hypothesize a change in executive function because these are thought to be abilities that solidify in early childhood (online Appendix D.8). As expected, there is no statistically significant change from treatment. This is one instance where the NSF proposal says otherwise, but our assessment of the literature changed between that proposal and data collection. Hence, we exclude executive function from the index. Including it would not materially change our conclusions, since the self-control measure performs weakest of all the noncognitive skills. Online Appendix D.8 reports these robustness checks.

Table 6—Impacts on Mental Health, Substance Abuse, and Quality of Social Networks

		ITT regression: ($N = 947$)											
		Therapy only				Cash only				Both			
				p-value				*p*-value				*p*-value	
Outcome, *z*-score	Control mean (1)	ITT (2)	SE (3)	Unadj. (4)	Adj. (5)	ITT (6)	SE (7)	Unadj. (8)	Adj. (9)	ITT (10)	SE (11)	Unadj. (12)	Adj. (13)
Panel A. 2–5 weeks													
Positive self-regard/ mental health	−0.168	0.111	[0.091]	0.222	0.681	0.137	[0.093]	0.139	0.582	0.343	[0.094]	**<0.001**	**0.004**
Positive self-regard	−0.179	0.121	[0.089]	0.178	0.419	0.151	[0.095]	0.111	0.338	0.346	[0.094]	**<0.001**	**0.001**
Depression and distress	0.095	−0.022	[0.099]	0.827	0.827	−0.063	[0.104]	0.546	0.766	−0.210	[0.101]	**0.039**	0.155
Substance abuse (0–3 index)	1.378	−0.249	[0.070]	**<0.001**	**0.006**	−0.060	[0.072]	0.407	0.833	−0.197	[0.071]	**0.006**	0.067
Usually drinks	0.671	−0.107	[0.041]	**0.009**	0.062	−0.023	[0.041]	0.575	0.928	−0.030	[0.040]	0.444	0.928
Usually uses marijuana	0.495	−0.111	[0.037]	**0.003**	**0.022**	−0.017	[0.036]	0.640	0.928	−0.141	[0.036]	**<0.001**	**0.001**
Usually takes hard drugs	0.212	−0.030	[0.027]	0.271	0.815	−0.020	[0.030]	0.496	0.928	−0.024	[0.030]	0.438	0.928
Quality of social network	−0.241	0.147	[0.062]	**0.018**	0.178	0.109	[0.061]	0.076	0.422	0.325	[0.062]	**<0.001**	**<0.001**
Peer quality	−0.160	0.207	[0.091]	**0.023**	0.190	0.014	[0.095]	0.881	0.996	0.235	[0.094]	**0.013**	0.120
Quality of family relations	−0.192	0.106	[0.099]	0.286	0.874	0.131	[0.105]	0.212	0.812	0.307	[0.099]	**0.002**	**0.024**
Ex-commander ties	−0.141	−0.011	[0.038]	0.776	0.996	0.013	[0.038]	0.730	0.996	−0.038	[0.038]	0.321	0.874
"Big men" ties	−0.012	0.015	[0.104]	0.889	0.996	0.039	[0.107]	0.714	0.996	0.172	[0.107]	0.109	0.596
Panel B. 12–13 months													
Positive self-regard/ mental health	−0.050	0.032	[0.091]	0.724	0.992	−0.034	[0.091]	0.707	0.992	0.205	[0.090]	**0.024**	0.287
Positive self-regard	−0.054	0.029	[0.093]	0.754	0.987	−0.035	[0.092]	0.709	0.987	0.222	[0.090]	**0.014**	0.072
Depression and distress	0.022	−0.031	[0.093]	0.741	0.987	0.021	[0.094]	0.824	0.987	−0.113	[0.096]	0.238	0.676
Substance abuse (0–3 index)	1.091	−0.065	[0.061]	0.282	0.951	0.063	[0.062]	0.305	0.952	−0.057	[0.060]	0.335	0.952
Usually drinks	0.766	−0.063	[0.042]	0.135	0.631	−0.047	[0.042]	0.267	0.813	−0.045	[0.043]	0.288	0.813
Usually uses marijuana	0.490	−0.024	[0.035]	0.502	0.924	0.018	[0.035]	0.610	0.935	−0.042	[0.036]	0.238	0.812
Usually takes hard drugs	0.196	−0.005	[0.031]	0.864	0.973	0.079	[0.033]	**0.016**	0.123	0.006	[0.031]	0.849	0.973
Quality of social network	0.066	0.063	[0.092]	0.496	0.971	−0.044	[0.092]	0.632	0.992	0.139	[0.095]	0.145	0.826
Peer quality	0.040	0.011	[0.088]	0.898	0.999	−0.070	[0.089]	0.430	0.982	0.017	[0.090]	0.852	0.999
Quality of family relations	−0.019	0.124	[0.099]	0.213	0.862	0.070	[0.100]	0.485	0.986	0.129	[0.097]	0.185	0.841
Ex-commander ties	0.176	0.004	[0.076]	0.953	0.999	0.026	[0.078]	0.740	0.998	−0.139	[0.074]	0.060	0.458
"Big men" ties	0.120	0.001	[0.155]	0.997	0.999	−0.130	[0.152]	0.395	0.979	0.071	[0.160]	0.657	0.998

Note: See Table 5.

control thoughts and actions.[47] Second, we used eight questions from the NEO-five factor personality inventory to assess conscientiousness (Costa and McCrae 1997). Topics included following societal rules and controlled, careful behavior. Third, we took seven questions on perseverance from the GRIT scale (Duckworth and Quinn 2009), which captures the ability to press on in the face of difficulty. Finally, we selected eight questions on reward responsiveness—whether they are motivated

[47] Examples include "I buy things on impulse" or "I say things without thinking."

by immediate, typically emotional rewards—from the Behavioral Inhibition/Behavioral Activation Scale.[48] We adapted the scales and questions to the context and Liberian English. Online Appendix D lists all questions.

In the short term, respondents reported little change in self-control skills. In fact, self-control was the only one of the six families that did not show a statistically significant increase in the short run, using unadjusted *p*-values. In contrast, after a year, both therapy and therapy plus cash are associated with increased self-control of 0.16 and 0.24 standard deviations. The latter is statistically significant at the 5 percent level with regular standard errors but not with the adjustment for multiple comparisons. Looking at the components of the self-control index, all point in the expected direction, though the magnitudes and precision are greatest with impulsivity and reward responsiveness.

Time preferences enter into our theoretical model differently than self-control, but an obvious question is whether they are distinct. The correlation between the self-control and time preference summary indices is 0.33, significant at the 1 percent level. If we combine the time preference and self-control measures into a single summary index, therapy alone and therapy plus cash both have statistically significant positive impacts of roughly 0.22 standard deviations after 2–5 weeks and 0.26 after 12–13 months (online Appendix D.8).

We must be cautious because all self-control scales are self-reported, and treated men could simply be repeating back their lessons. There is some evidence this is not so. We divide the 32 self-control questions into two indexes: questions with high (44 percent) and low (56 percent) emphasis in the curriculum.[49] Table 7 reports the ITT estimates after a year. The effect of cash and therapy is at least as large for low emphasis items.

Anticriminal Identity/Values.—Social identity and values are not straightforward to measure, and we are not aware of existing models. Based on our qualitative work, we developed three main measures, which we assemble into a single family index. First, we attempted to measure values directly, using a set of 33 self-reported attitudes toward the appropriate use crime and violence in the men's own lives—indicators of the degree to which they had internalized mainstream social norms.[50] Second, at the one-year endline we measured an index of prosocial behaviors, including group memberships, group and community leadership, and contributions to local public

[48] Examples include "I will often do things for no other reason than that they might be fun" or "When I see an opportunity for something I like I get excited right away." Previous research has linked disruptions in and extremes of reward motivation to substance abuse (Robinson and Berridge 2000).

[49] We rated each index component on a scale of 0 (not emphasized) to 4 (very emphasized). We then defined low-emphasis components as those rated 0 or 1 and high emphasis components as those rated 2 or above. These results are unchanged for using 1.5 or 2 as the emphasis cutoff.

[50] Our approach drew on the way social psychologists distinguish between social norms versus attitudes, but we focus on the attitudes alone. Social norms capture what people think others do or should do in a particular situation, and surveys sometimes then ask the same respondents what they believe is appropriate behavior (i.e., a norm deviation). Our questions are akin to these attitudinal questions. (We did not measure perceived norms because we did not have the space, nor did we deem the measure a priority.) Specifically, we asked 11 questions on attitudes to the use of violence to solve community or personal problems, such as mob killings of suspected thieves, or attacking their unfaithful wife's lover. We also asked 12 questions about their attitude toward participating in crime, including whether they would feel fine taking unwatched goods or stealing $100 from someone's pocket. We also asked about six hypothetical forms of political violence, including whether they discuss protesting with friends or making trouble or conflict with the authorities.

Table 7—Program Impacts on Noncognitive Skills and Values According to Their Coverage in the STYL Curriculum, 12–13 Month Only

Outcome (number of question in index), z-score	Control mean (1)	ITT regression: ($N = 947$ subjects) Therapy only ITT (2)	Therapy only SE (3)	Cash only ITT (4)	Cash only SE (5)	Both ITT (6)	Both SE (7)
Summary index of self-control skills							
Topics emphasized in curriculum (16)	−0.092	0.169	[0.090]	0.026	[0.091]	0.170	[0.093]
Topics not emphasized in curriculum (16)	−0.024	0.054	[0.098]	−0.070	[0.101]	0.232	[0.099]
Summary index of anticriminal/ antiviolent values[a]							
Topics emphasized in curriculum (8)	0.041	−0.085	[0.107]	0.012	[0.110]	−0.116	[0.108]
Topics not emphasized in curriculum (21)	0.070	−0.062	[0.087]	0.028	[0.087]	−0.184	[0.086]

Notes: The table reports intent to treat estimates of the effect of each treatment arm controlling for baseline covariates and block fixed effects. We have subdivided the summary indexes reported in Table 5 by their coverage of the specific topics in the STYL curriculum.

[a]These variables were not collected during every phase/round, so their regressions have a smaller sample size.

goods. (These are more of a behavior than a skill or preferences, but we hypothesized that it would be a reasonable proxy, allowing us to infer preferences from behaviors.) Finally, the therapy encouraged men to change their appearance as part of the identity change, and we asked survey enumerators to record their subjective impressions: quality of dress, shoes, cleanliness, and smell.[51]

In the short run, the family index improves by roughly 0.2 standard deviations in all treatment arms, significant at the 5 percent level with unadjusted p-values but not with adjusted p-values. With therapy followed by cash, the identity/values index rises 0.27 standard deviations, with $p = 0.067$ with our conservative adjustment. After a year, these treatment effects attenuate somewhat, particularly the change in appearance (which reverses in sign), and so the change is not statistically significant at conventional levels. Among the three components, however, the largest change appears to be in self-reported anticriminal and antiviolent values, and the impacts are sustained after a year in magnitude (though not statistically significant).[52]

Positive Self-Regard/Mental Health.—Half our mental health family index is positive self-regard. Poor self-regard has been linked with many aspects of negative behavior and counterproductive or extreme risk-seeking behavior (Coopersmith 1967). Some research (e.g., Judge et al. 2002) suggests self-regard is captured by an interrelated set of psychological scales, including: (i) neuroticism, the tendency to experience emotional instability or anxiety, assessed with eight questions from the NEO-5 factor personality inventory (Costa and McCrae 1997); (ii) self-esteem, assessed with eight questions such as, "I am able to do things as well as most other people" or "I take a positive attitude toward myself"; and (iii) locus of control, the

[51] Unlike the other two measures, we did not prespecify appearance as a reflection of identity and values in the NSF proposal. But it is difficult to see where else it belongs, and at the suggestion of referees we include it in this family.

[52] As with self-control, we divide the 29 value questions into 2 indexes by high and low emphasis in the curriculum. Table 7 reports the ITT estimates after a year. The effect of cash and therapy is at least as low for the low emphasis components.

extent to which individuals believe they versus fate control events affecting them, measured using eight questions from a standard questionnaire (Sapp and Harrod 1993). Arguably related to positive self-regard, we also collected a classic happiness measure, asking men to rank their subjective well-being in absolute terms and relative to others in their community.[53]

A second element of the mental health index is depression and distress. We assessed 6 symptoms of depression and 12 symptoms of posttraumatic stress, based on a locally adapted instrument used previously with ex-combatant populations in Liberia (Blattman and Annan 2016). We group this with positive self-regard as a mental health family in the interests of minimizing the number of families.

We treat positive self-regard/mental health as distinct from the anticriminal identity and values because, in principle, a positive self-image and criminal/outcast social identity are compatible. This simply happens to be uncommon in Liberia, where there is typically little social esteem associated with outcast/criminal social category. Moreover, the main purpose of the anticriminal measure was not to capture the quality of the men's self image, but rather their change in values, as illustrated in the theoretical model.

In the short run, a family index of these measures rises 0.34 standard deviations from therapy plus cash (significant at the 1 percent level using multiple comparisons). After a year, these treatment effects attenuate and are significant with unadjusted *p*-values only. Examining components suggests that the effects are driven most of all by the subjective well-being measure and self-esteem, although this does not stand up to adjusted *p*-values.

Substance Abuse.—Therapy tried to equip participants with strategies to cut back substance abuse, and while an obvious outcome of interest, it was not one we specified in advance of primary interest for two reasons: the program's overwhelming focus on antisocial behaviors, with drug use seen by NEPI as only a modest factor; and the fact that systematic reviews of CBT do not find support for its effectiveness treating substance use disorder (Hofmann et al. 2012). Nonetheless, if the therapy decreased substance use, we could see both economic and antisocial behavior change.

In the short run, reports of daily use of in the control group are 67 percent for alcohol, 50 percent for marijuana, and 21 percent for hard drugs. An index of all three indicators (0–3) fell 0.20, significant at the 10 percent level using adjusted *p*-values. At the one-year endline, reports of daily use of in the control group are roughly similar. An index of all three indicators (0–3) fell only 0.06 after a year as a result of therapy and cash (not statistically significant).

Quality of Social Networks.—Finally, we also assessed risky social networks.[54] We did not prespecify a change here, but over the course of the qualitative interviews,

[53] We asked about well-being, health, wealth, and power in absolute terms. We asked about wealth, respect, power, and access to services in relative terms. Each used a picture of a ladder with ten rungs. The summary index is the average of each ladder. Patterns are broadly similar across all ladders.

[54] In some settings, neighborhood changes would also indicate a change in risky behavior, but not in Liberia. Most high-crime neighborhoods in Monrovia are mixed-income residential and market areas where high-risk men

respondents repeatedly talked about changing peer groups to avoid crime, violence, and drugs. We thus measured the traits, positive and negative, of men's five closest peers.[55] We also asked about closeness to and support received from family members, former rebel commanders, and "big men" (intended to connote a criminal boss). A summary index of positive social networks increased in the short run by 0.15 standard deviations with therapy and 0.33 standard deviations with therapy plus cash (the latter significant at the 1 percent level with adjusted *p*-values). After a year, the point estimates remain positive, but are about one-half as large and not statistically significant.

D. *Insights from Qualitative Interviews and Observation*

One of the strongest impressions we gained from interviews was the importance men attached to identity change, or what NEPI called *transformation*. Nearly all of the subjects described feeling ostracized at baseline, and many reported that the therapy pushed them to believe they could be someone better for the first time. The facilitators played an important role here. The participants we interviewed unanimously had admiration and praise for the facilitators, highlighting that their backgrounds demanded respect and provided credibility, while their personal stories of change were encouraging. Beyond modeling the change in social identity, men reported the facilitators were also sometimes the first people to treat them with seriousness and respect, and this built their confidence to reintroduce themselves to community members, or to expose themselves to banks and shops.

Attempts to behave normally, especially the exposure to new social situations, seemed to reinforce skill and identity change. Many of the men failed in their plans, or experienced stigma in their shop or bank visits. In group sessions, men discussed what went wrong and why (such as poor decisions, or choice of dress). Men with setbacks learned from and were encouraged by the positive experiences of others. And facilitators sometimes observed men's homework attempts and coached them through difficulties.

Men's appearance also transformed during therapy. The first day men arrived with long or messy hair, facial hair, dirty or ripped clothing, wearing t-shirts with shorts and sandals. Their demeanor was tough, and their appearance signaled outcast status. Haircuts were offered in week two, and many men took advantage, symbolizing the change. Others showed up beforehand having gotten a haircut on their own. Similarly, before the unit on hygiene, some men began arriving in pants, shoes, and collared shirts. Typically a few men in each group resisted these changes. But seeing the positive experiences of others, they too began to arrive more clean cut, trying out the new identity. The survey results confirm a short-term change in appearance. The absence of 12–13-month change is puzzling.

---

are a small minority. They live on the margins, often in abandoned areas within these neighborhoods (building sites, forested groves, garbage dumps, and abandoned buildings). Men who turn their lives around do not need to move neighborhoods, but rather stay where they are but move in different, more mainstream social circles, avoiding high-risk hangouts. Thus, we do not report on neighborhood movements.

[55] We ask men who their five closest peers are, by name, and then ask whether they hold any of 20 qualities ranging from positive (they work hard, save, go to school) or negative (they steal, do drugs, get in fights).

A year later, therapy participants also described applying skills of self-regulation in their lives. To avoid fights, they used new tactics: removing themselves from emotionally charged situations, allowing space to process their feelings, and ignoring negative automatic thoughts in the favor of more controlled thinking. Related were improved social and communication skills. Interviewees described how such skills allowed them to engage with community members or in disputes and express themselves without anger or violence.

Not only did the community regard them differently, many said, but troubled young men began coming to them for advice and lessons learned from the therapy once they saw the sudden and sustained change—another important source of reinforcement, and perhaps one reason we do not see a change in peer quality in the data.

VI. Can We Believe Our Self-Reported Data?

Self-reported data raise several worries, the most serious being measurement error correlated with treatment. For instance, men who receive an antiviolence intervention might be more likely to tell us they are nonviolent, overestimating the estimated treatment effect of therapy.

This kind of bias is hard to square with the patterns of effects we observed. Therapy followed by cash would have to induce systematic errors where therapy or cash alone did not. Nonetheless, this is possible—for example if the largest misreporting were associated with larger past benefits. Thus, concerned that our survey measure, y^s, may be biased, we set out to intensively validate some measures, y^v. If y^v is closer to the true behavior, y^*, this allows us to estimate the degree and direction of bias. We summarize the approach, empirical strategy, and results here, with details in online Appendix F and Blattman et al. (2016).

A. *Approach to Validation*

Of more than 4,000 endline surveys, we randomly selected 7.3 percent and retested answers to six survey-based measures with two-week recall periods. We chose four potentially sensitive behaviors—marijuana use, thievery, gambling, and homelessness. We also chose two everyday expenditures that could be subject to recall bias or other error—paying to watch television in a video club, and paying to charge a mobile phone. We chose these six because we wanted a diverse set of behaviors with similar recall periods. We also wanted very specific behaviors (e.g., stealing rather than any crime, or marijuana rather than substance abuse). Finally, we wanted outcomes that were a primary focus of the treatment (e.g., stealing) and others that were not (gambling or expenditures).

We used intense qualitative work—in-depth participant observation, open-ended questioning, and efforts to build relationships and trust—to try to elicit more truthful answers. Over several days of trust-building and conversation, plus direct observation, we tried to elicit a direct admission or discussion of the behavior.

We selected and trained eight of the study's most talented qualitative research staff as validators, all Liberians. In the ten days following the survey, a validator visited the respondent over four days, spending several hours each day in conversation and observation. Validators shadowed respondents as they went about their day,

rather than conduct formal interviews. They raised target topics through indirect questions while chatting.

Validators developed techniques to foster trusting relationships and to build rapport: becoming close to street leaders; eating meals with subjects; sharing personal information (including similar acts they or their friends engaged in); and mirroring participants' appearance and vernacular as appropriate. Validators would also observe the respondent's behavior from afar, as well as converse with peers and family. The goal was to attain insider status, and thus reduce the chance of misreporting. The premise was that time, a focus on a small number of behaviors, and trust/rapport building would mean that respondents were less willing, or feel less able, to deceive a more familiar person, who also knows them better. Validators also had the opportunity to clear up misunderstandings and get a more accurate assessment of the behaviors. By discussing sensitive behaviors openly, relating their own experiences and that of friends, validators sought to dispel any notion that certain answers are more desirable, or would result in any strategic gains.

Without knowing the respondent's survey response, y^s, the validators coded an indicator of whether or not the respondent engaged in the behaviors in the two weeks prior to the survey, y^v. The authors reviewed the evidence and the coding for every case. In general, we used a relatively high standard of evidence, only coding $y^v = 1$ for a direct admission of the behavior or persuasive statements that they did not engage in the behavior.[56]

If this technique simply reproduced the errors in the survey data, then the validation is little help. The key assumption is that four days of building trust and gathering extensive information, regarding just six behaviors, reduced experimenter demand and other biases correlated with treatment compared to responses during a 300-question, 90-minute questionnaire.

Nonetheless, y^v is not free from error. Online Appendix F.1 reviews our approach and its limitations in more detail. Many of these limitations—the requirement of a direct admission, the disruption in people's lives, errors in recall periods, or increased social desirability bias from scrutiny—undoubtedly led to systematic errors in y^v. These errors, however, are not necessarily correlated with treatment. This is possible, for example, because validators could have learned men's treatment status in conversation, and this could have biased their coding. Nonetheless, we designed the trust-building and evidentiary standards to minimize this risk.

B. *Survey-Validation Differences*

Of the 297 men we selected for validation, we found and validated 240 (81 percent).[57] Table 8 reports the means of y^s and y^v in the full sample and each treatment arm, as well as the percentage of times the two measures agree; y^s and y^v are

[56] The validators only witnessed or received third-party evidence of the behavior in one-fifth of cases, but neither was considered sufficient evidence for a final coding. Both had to be followed by questions confirming that the respondent also engaged in the behavior in the two weeks prior to the survey. In general, we used a relatively high standard of evidence, only coding $y^v = 1$ if the validator directly observed the behavior or the respondent directly admitted it.

[57] Attrition was higher than the survey as we could not validate the behaviors of men who migrated across the country. Attrition was not correlated with treatment or baseline covariates (Blattman et al. 2016).

Table 8—Comparison of Survey and Qualitative Validation Means at Endline

	Potentially sensitive behaviors					Expenditures			
	All (0–4) (1)	Steal (2)	Marijuana (3)	Gamble (4)	Homeless (5)	All (0–2) (6)	Video (7)	Phone (8)	All (0–6) (9)
Panel A. Full sample									
Survey mean	1.12	0.22	0.48	0.18	0.23	0.82	0.42	0.39	1.93
(y^s)	(1.14)	(0.42)	(0.50)	(0.39)	(0.42)	(0.73)	(0.50)	(0.49)	(1.31)
Validation mean	1.21	0.20	0.51	0.29	0.21	1.09	0.61	0.48	2.30
(y^v)	(1.18)	(0.40)	(0.50)	(0.45)	(0.41)	(0.74)	(0.49)	(0.50)	(1.21)
Percent in agreement		79	85	72	82		62	82	
Panel B. Control group									
Survey mean	1.25	0.27	0.48	0.23	0.27	0.68	0.37	0.32	1.93
(y^s)	(1.31)	(0.45)	(0.50)	(0.43)	(0.45)	(0.70)	(0.49)	(0.47)	(1.44)
Validation mean	1.30	0.23	0.49	0.34	0.23	1.18	0.65	0.54	2.48
(y^v)	(1.23)	(0.42)	(0.50)	(0.48)	(0.42)	(0.70)	(0.48)	(0.50)	(1.21)
Percent in agreement		80	88	72	77		47	75	
Panel C. Therapy only									
Survey mean	1.06	0.19	0.48	0.17	0.22	0.81	0.41	0.41	1.87
(y^s)	(1.11)	(0.39)	(0.50)	(0.38)	(0.42)	(0.75)	(0.50)	(0.50)	(1.35)
Validation mean	1.09	0.17	0.48	0.24	0.20	0.98	0.54	0.44	2.07
(y^v)	(1.14)	(0.38)	(0.50)	(0.43)	(0.41)	(0.76)	(0.50)	(0.50)	(1.24)
Percent in agreement		80	89	74	80		72	81	
Panel D. Cash only									
Survey mean	1.03	0.21	0.49	0.13	0.21	0.77	0.37	0.40	1.81
(y^s)	(1.16)	(0.41)	(0.50)	(0.34)	(0.41)	(0.71)	(0.49)	(0.49)	(1.35)
Validation mean	1.32	0.23	0.53	0.33	0.24	1.00	0.55	0.45	2.32
(y^v)	(1.26)	(0.42)	(0.50)	(0.47)	(0.43)	(0.81)	(0.50)	(0.50)	(1.33)
Percent in agreement		76	82	74	90		56	85	
Panel E. Therapy + cash									
Survey mean	1.13	0.22	0.48	0.21	0.22	0.98	0.54	0.44	2.11
(y^s)	(0.98)	(0.42)	(0.50)	(0.41)	(0.42)	(0.73)	(0.50)	(0.50)	(1.11)
Validation mean	1.11	0.19	0.52	0.24	0.16	1.17	0.70	0.48	2.29
(y^v)	(1.11)	(0.40)	(0.50)	(0.43)	(0.37)	(0.68)	(0.46)	(0.50)	(1.05)
Percent in agreement		81	83	68	81		71	87	
Observations	239	238	238	238	239	239	238	239	239

Notes: The table reports the means (standard deviations) of the survey and the qualitatively validated measures for the full sample and by treatment arm. Percent in agreement is the percentage of respondents for whom the survey indicator equals the qualitatively validated indicator.

identical about 80 percent of the time for sensitive measures and about 70 percent of the time for expenditures. As expected, however, $\bar{y}^s < \bar{y}^v$: the average person reported 1.21 sensitive behaviors and 1.09 expenditures in validation, and 1.12 sensitive behaviors and 0.82 expenditures in the survey.

With this sample, only the underreporting of expenditures is statistically significant. We report *t*-tests of the simple difference, $y_i^s - y_i^v$, in online Appendix F.2.1, as well as a discussion of patterns of under- and overreporting. Expenditure underreporting appears to be largest in the control group, possibly because they are trying to appear more needy. Among sensitive behaviors, underreporting is generally less than 10 percent of the survey means, and is only statistically significant in the case

of gambling. This is mainly driven by the cash only arm, who may have been reluctant to report spending the grant this way.

C. *Is Measurement Error Correlated with Treatment?*

Empirical Strategy.—If we believe that the validation measure is closer to the true behavior, then one way to test for bias in the survey-based treatment effects is to take the difference $y_i^s - y_i^v$, our proxy of measurement error for person i, and regress it on treatment:

$$y_i^s - y_i^v = \beta_0 + \beta_1 T_i + \mu_i. \tag{2}$$

If $\beta_1 < 0$ for sensitive measures, then treated men were more likely to underreport bad behaviors, and our survey-based treatment effects may overestimate the decline in antisocial behaviors. And if $\beta_1 > 0$ for expenditures, then treated men may have overreported their expenditures more than the control group, and our survey-based treatment effects may overestimate the short-run increase in income.

With a sample of 240, we estimate we are powered to detect average under- or overreporting of at least 14 percent, and error correlated with treatment of 28 percent.[58] Because of power concerns, we pay close attention to the sign, magnitude, and confidence interval for β_1.

Of course, the crucial assumption is that y^v is closer to the true behavior. This parallels the *no liars* and *no design effects* assumptions in list experiments. The assumption cannot be tested directly, but can only be argued on context and the quality of the approach.

We can also let misreporting vary by whether validation confirmed the behavior:

$$y_i^s = \tilde{\beta}_0 + \tilde{\beta}_1 T_i + \tilde{\beta}_2 y_i^v + \tilde{\beta}_3 (y_i^v \times T_i) + \tilde{\mu}_i. \tag{3}$$

Equation (2) is simply the special case where $\tilde{\beta}_2 = 1$ and $\tilde{\beta}_3 = 0$.[59] We are mainly interested in whether $\tilde{\beta}_1 \neq 0$ and $\tilde{\beta}_1 + \tilde{\beta}_3 \neq 0$. The disadvantage of this more flexible form is statistical power, especially with three treatment arms.[60] We are also interested in correcting for the average bias in survey-based treatment effects, which we get from β_1 from equation (2). But the more flexible form provides insight into

[58] Our target sample of 297 was the maximum number of interviews we felt qualified validators could manage logistically. We calculated minimum detectable effects (MDE) using a two-sided hypothesis test with 80 percent power at a 0.05 significance level, using baseline and block controls when calculating the R^2 statistic. We calculated an MDE for both the 0–2 expenditures index and the 0–4 sensitive behaviors index. The expenditures index had a mean of 0.82 in the survey and an MDE of 0.13 for general over- and underreporting, and 0.29 for a treatment effect on misreporting. The sensitive behaviors index had a mean of 1.12 in the survey and an MDE of 0.2 for general over- and underreporting, and 0.36 for any treatment effect on misreporting. We estimate that doubling the sample size would have increased power by about one-third.

[59] Online Appendix F.1 derives and interprets these regressions in more detail.

[60] With 240 observations in total, each parameter is estimated off of roughly 30 observations, putting us on a steep part of the power curve.

the patterns of measurement error. For instance, if underreporting is concentrated among men who commit crimes and were treated, then $\tilde{\beta}_1 + \tilde{\beta}_3 < 0$.[61]

Results for Sensitive Behaviors.—We estimate equations (2) and (3) in Table 9, including block fixed effects.[62] For sensitive behaviors, almost none of the coefficients on treatment indicators or interactions are statistically significant. We see little evidence of the therapy inducing a desirability bias, and indeed the effects run in the opposite direction.

Indeed, looking at the index of four sensitive measures (panel A, column 5), β_1 is actually greater than zero for therapy plus cash, implying that the impacts of therapy plus cash are, if anything, larger than the survey data imply. Online Appendix F.3 displays these updated treatment effects. For example, using survey data alone, the treatment effect (standard error) of therapy and cash on the sensitive behaviors index is -0.4 (0.09), a 36 percent decrease. The results from panel A, column 5 suggest that the adjusted treatment effect should be -0.516 (0.194), significant at the 1 percent level.

The results of the more flexible regression in panel B, column 5 show that these averages conceal important heterogeneity. Treated men who we think did not engage in the sensitive behaviors tend to overreport them ($\tilde{\beta}_1^{Both} > 0$), and treated men engaged in the sensitive behaviors seem to underreport them ($\tilde{\beta}_1^{Both} + \tilde{\beta}_3^{Both} < 0$).

Results for Expenditures.—All treatment arms associated with a roughly 0.3 increase in our proxy for measurement error (panel A, column 8). There is underreporting across all arms, but it is greatest in the control group. This could have implications for one of our main findings, on income. Using survey data, the treatment effect of cash only on the 2-item expenditure index is 0.08 (0.052), which is consistent with the short-run increase in consumption we observed among cash recipients. But adjusting for observed measurement error, the adjusted treatment effect is -0.205 (0.143).

Interpretation.—Our qualitative work suggests two explanations. The men have been members of a subculture where drugs, crime, and gambling are commonplace, and admitting to the behaviors in a survey carries little stigma. Speculatively, therapy may have accustomed men to talking about these behaviors or reduced stigma. As for expenditures, control men may have acted strategically, trying to appear poorer in the hopes they would be eligible for assistance. We discuss implications for our conclusions in the following section.

[61] If we exclude the block fixed effects used for estimating ITT effect, as in equation (1), then $\tilde{\beta}_0$ also contains information: if men honestly report crime in the survey then $\tilde{\beta}_0$ should be close to 0 and $\tilde{\beta}_2$ should be close to 1, while if there is a general desirability bias in the survey, then $\tilde{\beta}_0 + \tilde{\beta}_2 < 1$. See online Appendix F.2.2 for this analysis. In general, estimates of $\tilde{\beta}_0$ suggest that sensitive behaviors are 12–15 percent more likely to be reported in the survey, possibly because of the validator's fidelity to the two-week recall period and specific definitions, or a general conservatism, but there is no evidence this *survey overreporting* is correlated with treatment, which is the main purpose of the analysis.

[62] That is, in equations (2) and (3) we actually estimate β_{0j} and $\tilde{\beta}_{0j}$, which is necessary to identify treatment effects when the probability of treatment assignment varies by block. The results without block fixed effects (not shown) are qualitatively similar.

TABLE 9—ESTIMATES OF THE CORRELATION BETWEEN TREATMENT AND MEASUREMENT ERROR

	Dependent variable: ($N = 239$)							
	$y^s - y^v$, sensitive behaviors					$y^s - y^v$, expenditures		
Covariate	Stealing (1)	Marijuana (2)	Gambling (3)	Homeless (4)	All (0–4) (5)	Video club (6)	Phone charging (7)	All (0–2) (8)
Panel A. Constrained, with block fixed effects (equation (2))								
β_0 (constant)	−0.029	0.062	−0.109	0.093	0.015	−0.326	−0.194	−0.517
	[0.087]	[0.061]	[0.093]	[0.076]	[0.177]	[0.118]	[0.095]	[0.158]
β_1								
Therapy	−0.019	0.015	0.025	−0.025	−0.004	0.170	0.174	0.339
	[0.084]	[0.057]	[0.097]	[0.091]	[0.199]	[0.102]	[0.085]	[0.132]
Cash	−0.038	−0.042	−0.085	−0.077	−0.237	0.109	0.165	0.269
	[0.088]	[0.067]	[0.090]	[0.079]	[0.195]	[0.111]	[0.078]	[0.134]
Both	−0.006	−0.024	0.077	0.031	0.079	0.127	0.181	0.304
	[0.080]	[0.062]	[0.095]	[0.089]	[0.183]	[0.103]	[0.075]	[0.115]
	Dependent variable: ($N = 239$)							
	y^s, sensitive behaviors					y^s, expenditures		
Covariate	Stealing (1)	Marijuana (2)	Gambling (3)	Homeless (4)	All (0–4) (5)	Video club (6)	Phone charging (7)	All (0–2) (8)
Panel B. Unconstrained, with block fixed effects (equation (3))								
$\tilde{\beta}_0$ (constant)	0.301	0.098	0.231	0.283	0.976	0.367	0.047	0.048
	[0.140]	[0.092]	[0.118]	[0.129]	[0.287]	[0.130]	[0.093]	[0.208]
$\tilde{\beta}_1$								
Therapy	−0.022	0.010	−0.009	−0.036	0.154	−0.190	0.100	−0.011
	[0.070]	[0.077]	[0.074]	[0.079]	[0.228]	[0.124]	[0.072]	[0.207]
Cash	0.003	0.025	−0.079	−0.138	−0.069	−0.072	0.040	0.089
	[0.068]	[0.081]	[0.064]	[0.070]	[0.220]	[0.139]	[0.068]	[0.219]
Both	−0.013	0.025	0.064	−0.025	0.271	−0.113	0.041	−0.150
	[0.064]	[0.081]	[0.076]	[0.083]	[0.241]	[0.138]	[0.064]	[0.209]
$\tilde{\beta}_2(y^v)$	0.496	0.735	0.315	0.405	0.677	0.038	0.504	0.328
	[0.158]	[0.096]	[0.108]	[0.158]	[0.108]	[0.123]	[0.096]	[0.129]
$\tilde{\beta}_3$								
Therapy × y^v	−0.166	−0.014	−0.131	−0.020	−0.210	0.434	0.079	0.222
	[0.234]	[0.125]	[0.176]	[0.220]	[0.147]	[0.169]	[0.143]	[0.169]
Cash × y^v	−0.232	−0.114	−0.063	0.286	−0.134	0.133	0.196	0.064
	[0.208]	[0.134]	[0.148]	[0.202]	[0.144]	[0.174]	[0.137]	[0.169]
Both × y^v	−0.064	−0.085	−0.241	0.066	−0.230	0.386	0.234	0.379
	[0.214]	[0.129]	[0.173]	[0.232]	[0.137]	[0.171]	[0.130]	[0.168]
p-value from test that $\tilde{\beta}_1 = \tilde{\beta}_3 = 0$								
Therapy	0.671	0.991	0.682	0.867	0.345	0.035	0.161	0.098
Cash	0.511	0.648	0.299	0.115	0.307	0.744	0.134	0.380
Both	0.908	0.784	0.369	0.941	0.240	0.023	0.054	0.005

Notes: The table reports the degree and direction of bias in our treatment effects. In panel A, we assume that our measurement error does not vary by whether or not the individual engages in the behavior, which allows for a simple way to use β_2 to adjust our ITT estimates. In panel B, we relax this assumption and let the measurement error vary by behavior and treatment arm at the cost of reduced statistical power.

VII. Discussion

A. *Lessons from the Cash Transfer*

One lesson is that these supposedly undisciplined men largely invested and saved a grant. Even accounting for the underreporting we see in gambling and other expenditures, little of the grants seem to have been spent on temptation goods. While

Evans and Popova (2017) see the same result in 19 other cash transfer programs, it's striking to see the same with this extreme group. Caution is also warranted, because of the evidence that the control group underreported expenditures. But in the short run, men seem to have used the cash for petty trade, earning returns to capital of at least 26 percent.[63]

There is only weak evidence, however, that criminal activities fell as business income rose. Those who received cash reduced their antisocial behaviors only 0.08 standard deviations (not significant) but reduced stealing by about 30 percent (significant at the 10 percent level with unadjusted *p*-values only). The direction of effects, however, is consistent with rural ex-combatants in Liberia, who shifted away from illicit activities when a much more intensive employment program raised their farm productivity (Blattman and Annan 2016).

Any investments and income gains disappeared within a year, however, in part due to poor property rights protections. The men's homes and neighborhoods were highly insecure. Extrapolating from reports of burglary and theft at each endline (from Table 3), men in our sample experienced a theft or robbery roughly eight times in the year after the grant. While treated men were no more likely to experience theft, they had more to lose, especially their savings and investment in nascent businesses.

Nonetheless, the fact that cash was well used is important, since concerns about temptation spending restrain political support for cash-based welfare programs. The men received a few months' worth of income, and basic consumption—especially basic shelter and food—improved for about that length of time. This is important.

Future research should study how to sustain the economic effects of cash. It may be that helping people relocate to better quality neighborhoods or enhance their personal security, or providing the information and means to gain necessary licenses or protection from security forces, would reduce expropriation. Alternately, programs can try to provide crude insurance. It is possible that regular cash transfers would stimulate enterprise development more than the one-time transfer we study (Bianchi and Bobba 2013; Karlan, Knight, and Udry 2015).

B. *Lessons from Behavior Change*

The interventions have extremely similar impacts on both antisocial behaviors and an index of all noncognitive skill and preference measures, suggesting that noncognitive change, broadly speaking, was a major source of behavior change. The impact of cash plus therapy on antisocial behaviors (0.31 standard deviations in two to five weeks and 0.25 after a year) is mirrored by a change of similar magnitude in noncognitive skills and preferences (0.43 standard deviations in two to five weeks and 0.25 after a year: see Figure 1). Likewise, therapy alone significantly affects antisocial behaviors and noncognitive changes within two to five weeks but not after

[63] For instance, the impact on earnings ($8.25 a month) represents a monthly return of 4.1 percent on the $200 grant, while the impact on nondurable consumption ($48 per month) represents a monthly return of 24 percent. While there are reasons these figures might overstate returns, recall that men only invested about 60 percent in the month after the grant, implying returns on actual investment are probably higher. Unfortunately we cannot say whether the cash grant passed a cost-benefit test in private monetary returns alone.

a year, and cash alone does not have a significant effect on either outcome in any period.

Among the six noncognitive families we defined, all but self-control skills showed a large and (with unadjusted standard errors) statistically significant decline after two to five weeks as a result of therapy and cash. Forward-looking time preferences, mental health (particularly positive self-regard), social networks (particularly peer quality and family relationships), and identity/values all show large and robust changes after adjusting for multiple comparisons. After a year, it is difficult to single out any one noncognitive skill or preference change as robust. Individually, the largest and most precise changes are in forward-looking time preferences and self-control skills.

Nonetheless, if we were to merge time preferences and self-control skills into a single index of *future orientation*, we see some evidence of sustained impacts as a result of therapy plus cash. The combined index increased by 0.22 standard deviations after two to five weeks (adjusted $p = 0.029$), and by 0.26 standard deviations after one year (adjusted $p = 0.068$). We must take this result with some caution, partly because we prespecified them as distinct measures, and partly because we do not see robust changes in short-term self-control. Nonetheless, a change in future orientation would echo the effects of adolescent CBT programs in Chicago that target similar automatic behaviors (Heller et al. 2015).

We see less conclusive evidence on the least standard aspect of the therapy—the focus on changing social identity and values. This family index increased 0.27 standard deviations after two to five weeks, significant at the 1 percent level with conventional standard errors but an adjusted $p = 0.067$. This index change driven by significant changes in both appearance and anticriminal values. But effects moderate after a year, particularly for appearance (and with the addition of prosocial behaviors), and become less precise.

These estimates contrast with men's qualitative personal accounts, where identity change was paramount. Qualitatively, the changes in appearance, in community regard, and the exposure to new places and situations seem to have been particularly important. So was the identity of the NEPI facilitators, and the fact that they modeled this identity change. This change has a basis in the theory underlying CBT: positive interactions challenged respondents' negative beliefs about themselves, and reinforced their identity as more responsible, mainstream members of society. Possibly identity and values are difficult to measure, and so this remains an important area for further innovation and testing.

In psychology, efficacy trials such as this one are typically followed by further trials that try to identify the "active ingredients," by varying modules and methods. This, plus more investment in measurement, seems like a fruitful area for research.

C. *Understanding the Cash–Therapy Interaction*

We did not expect that the effects of therapy would persist only when cash was received as well. Our theory predicted that the two interventions should have a larger effect only if cash raises earnings permanently, which was not the case.

Our qualitative evidence and psychological theory, however, suggest a hypothesis for testing in future trials: that receiving cash was akin to an extension of therapy, in that it provided more time for the men to practice independently and to reinforce

their changed skills, identity, and behaviors. The therapy was brief—just eight weeks long. It helped men change their intentions, identity, and behavior, and provided almost daily commitment and reinforcement. After eight weeks the men who received therapy alone had to contend with their usual economic and peer pressures. The grant, however, provided some men with the cash they needed to maintain their new identity—to avoid homelessness, to feed themselves, and to continue to dress well. They had no immediate financial need to return to crime.

The men could also do something consistent with their new identity and skills: execute plans for a business. This was a source of practice and reinforcement of their newfound skills and identity. It was also a form of performance, to themselves as well as their family and neighbors, who could see the men engage in legitimate business. Our qualitative interviews also suggested that the cash helped men to survive shocks. In this way, the grant may have parallels to *booster sessions* commonly used in therapy. A small body of experimental research on CBT for aggression or substance abuse indicates that follow-up therapy sessions weeks or months after the intervention improve 12–13-month outcomes (e.g., Lochman 1992).

Caution is warranted. We cannot reject the hypothesis, for instance, that positive reinforcement from winning a grant was enough to reinforce therapy. In future research, a comparison of extended therapy to shorter therapy plus cash would offer a more direct test.

Nonetheless, high short-run returns to capital and sustained social spillovers suggest that the combination of cash and therapy had promising returns. Since the private returns to the grant were temporary, however, the cost effectiveness rides mostly on the social benefits from roughly one fewer crime per week per person. These social returns are unknown. If these social returns are greater than $20 or $25 per crime, however, the STYL program is a promising investment on basis of crime reduction alone.

D. *Generalizability*

For several reasons this approach has promise beyond Liberia. First, the therapy was adapted from US-based CBT programs, suggesting that adaptability to other contexts is feasible. Second, we kept the intervention low-cost and created a publicly available manual, curriculum, and training guidelines to ease adaptation and replication.[64] Third, with time it should be possible to develop qualified and effective facilitators in other countries, not least because there are established methods for training counselors in CBT; general levels of education (and the number of social workers) are greater in most other countries; and new facilitators should emerge among graduates of the program, as with STYL.

The theory and results are also strikingly consistent with comparable US programs and best practice. The attention to noncognitive skill change and social identity, the targeting of the highest-risk men, as well as the nonresidential nature of the therapy, correspond closely to best practice in criminal rehabilitation in US correctional institutions (Andrews et al. 1990; Lipsey 2009). The 40–50 percent falls

[64] This manual can be found at http://chrisblattman.com/documents/policy/2015.STYL.Program.Manual.pdf.

in antisocial behaviors we observe are similar in proportion to the falls in arrests documented in Tennessee and Chicago (Little, Robinson, and Burnette 1994; Heller et al. 2015). Moreover, as in Chicago, the effects of therapy alone were temporary.

Other US work suggests that employment can be complementary to social and emotional counseling (Heller 2014). In low-income countries, however, where most employment programs will involve self-employment, property security and risk are important scope considerations. Cash transfers in other poor countries have generally led to higher and more persistent incomes, in part because the gains are not stolen. So the STYL program could arguably work even better in places with more secure property rights.

There are limits to generalizability of course. For instance, there were no gangs or armed groups vying for men in our sample. CBT-based approaches may be most effective against disorganized, impulsive crime and violence rather than organized crime. There is also selection onto the street, and a country which has experienced many negative shocks (such as Liberia) might have more high-potential young men who need only a little help to regress to the mean. On the other hand, our evidence from dropouts suggests that the most antisocial men stay, and the program is most effective with them. These limits are speculative without further testing, however, and replication and experimentation seem more than warranted given the results of these efficacy trials in Liberia, Chicago, and elsewhere.

REFERENCES

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–53.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz. 2011. "Personality Psychology and Economics." In *Handbook of the Economics of Education,* Vol. 4, edited by E. Hanushek, S. Machin, and L. Woessman, 1–181. Amsterdam: Elsevier.

Anderson, Michael L. 2008. "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* 103 (484): 1481–95.

Andrews, D. A., I. Zinger, R. D. Hoge, J. Bonta, P. Gendreau, and F. T. Cullen. 1990. "Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-analysis." *Criminology* 28 (3): 369–404.

Annan, J., C. Blattman, and R. Littman. 2015. "Mental Health Impacts of an Excombatant Reintegration Program." Unpublished.

Beck, Judith S. 2011. *Cognitive Behavior Therapy: Basics and Beyond.* New York: Guilford Press.

Becker, Anke, Thomas Deckers, Thomas Dohmen, Armin Falk, and Fabian Kosse. 2012. "The Relationship between Economic Preferences and Psychological Personality Measures." *Annual Review of Economics* 4 (1): 453–78.

Becker, G. S. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76: 169–217.

Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112 (4): 848–86.

Bianchi, Milo, and Matteo Bobba. 2013. "Liquidity, Risk, and Occupational Choices." *Review of Economic Studies* 80 (2): 491–511.

Blattman, Christopher, and Jeannie Annan. 2016. "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State." *American Political Science Review* 110 (1): 1–17.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics* 129 (2): 697–752.

Blattman, Christopher, Julian Jamison, Tricia Koroknay-Palicz, Katherine Rodrigues, and Margaret Sheridan. 2016. "Measuring the Measurement Error: A Method to Qualitatively Validate Survey Data." *Journal of Development Economics* 120: 99–112.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2012. "Experimentally Testing the Roots of Poverty and Violence: Changing Preferences, Behaviors, and Outcomes." http://chrisblattman.com/documents/research/2012.01.13_STYL_NSF_proposal.pdf.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia: Dataset." *American Economic Review.* https://doi.org/10.1257/aer.20150503.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel. 2008. "Symposium on Noncognitive Skills and Their Development: The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43 (4): 972–1059.

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics* 127 (4): 1755–812.

Christensen, Maya M., and Mats Utas. 2008. "Mercenaries of Democracy: The 'Politricks' of Remobilized Combatants in the 2007 General Elections, Sierra Leone." *African Affairs* 107 (429): 515–39.

Coopersmith, Stanley. 1967. *The Antecedents of Self-Esteem.* San Francisco: W. H. Freeman.

Costa, P. T., and R. R. McCrae. 1997. "Stability and Change in Personality Assessment: The Revised NEO Personality Inventory in the Year 2000." *Journal of Personality Assessment* 68 (1): 86–94.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

Del Vecchio, T., and K. D. O'Leary. 2004. "Effectiveness of Anger Treatments for Specific Anger Problems: A Meta-Analytic Review." *Clinical Psychology Review* 24 (1): 15–34.

Draca, Mirko, and Stephen Machin. 2015. "Crime and Economic Incentives." *Annual Review of Economics* 7: 389–408.

Duckworth, Angela Lee, and Patrick D. Quinn. 2009. "Development and Validation of the Short Grit Scale (Grit-S)." *Journal of Personality Assessment* 91 (2): 166–74.

Duckworth, A. L., and R. Shulze. 2011. "A Meta-Analysis of Convergent Validity Evidence for Self-Control Measures." *Journal of Research in Personality* 45 (3): 259–68.

Evans, David K., and Anna Popova. 2017. "Cash Transfers and Temptation Goods." *Economic Development and Cultural Change* 65 (2): 189–221.

Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2015. "Sex Workers, Self-Image and Stigma: Evidence from Kolkata Brothels." Unpublished.

Gottfredson, Michael R., and Travis Hirschi. 1990. *A General Theory of Crime.* Palo Alto: Stanford University Press.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *Quarterly Journal of Economics* 131 (4): 1973–2042.

Heckman, James J., and Tim Kautz. 2014. "Fostering and Measuring Skills: Interventions that Improve Character and Cognition." In *The Myth of Achievement Tests: The GED and the Role of Character in American Life,* edited by James Heckman, J. E. Humphries, and Tim Kautz, 341–430. Chicago: University of Chicago Press.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–82.

Heller, Sara B. 2014. "Summer Jobs Reduce Violence among Disadvantaged Youth." *Science* 346 (6214): 1219–23.

Heller, Sara B., Harold A. Pollack, Roseanna Ander, and Jens Ludwig. 2013. "Preventing Youth Violence and Dropout: A Randomized Field Experiment." National Bureau of Economic Research Working Paper 19014.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. 2015. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." National Bureau of Economic Research Working Paper 21178.

Hill, Patrick L., Brent W. Roberts, Jeffrey T. Grogger, Jonathan Guryan, and Karen Sixkiller. 2011. "Decreasing Delinquency, Criminal Behavior, and Recidivism by Intervening on Psychological Factors Other than Cognitive Ability: A Review of the Intervention Literature." National Bureau of Economic Research Working Paper 16698.

Hofmann, S. G., A. Asnaani, I. J. Vonk, A. T. Sawyer, and A. Fang. 2012. "The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-Analysis." *Cognitive Therapy and Research* 36 (5): 427–40.

Jolls, Christine, Cass R. Sunstein, and Richard Thaler. 1998. "A Behavioral Approach to Law and Economics." *Stanford Law Review* 50 (5): 1471–1550.

Judge, Timothy, Amir Erez, Joyce E. Bono, and Carl J. Thoresen. 2002. "Are Measures of Self-Esteem, Neuroticism, Locus of Control, and Generalized Self-Efficacy Indicators of a Common Core Construct?" *Journal of Personality and Social Psychology* 83 (3): 693–710.

Karlan, Dean, Ryan Knight, and Christopher Udry. 2015. "Consulting and Capital Experiments with Microenterprise Tailors in Ghana." *Journal of Economic Behavior and Organization* 118: 281–302.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.

Langberg, J. M., S. P. Becker, J. N. Epstein, A. J. Vaughn, and E. Girio-Herrera. 2013. "Predictors of Response and Mechanisms of Change in an Organizational Skills Intervention for Students with ADHD." *Journal of Child and Family Studies* 22 (6): 1000–12.

Lipsey, Mark W. 2009. "The Primary Factors that Characterize Effective Interventions with Juvenile Offenders: A Meta-Analytic Overview." *Victims and Offenders* 4: 124-47.

Little, G., K. Robinson, and K. Burnette. 1994. "Treating Offenders with Cognitive Behavioral Therapy: 5-Year Recidivism Outcome Data on MRT." *Cognitive-Behavioral Treatment Review* 3: 1–3.

Lochman, J. E. 1992. "Cognitive-Behavioral Intervention with Aggressive Boys: Three-Year Follow-Up and Preventive Effects." *Journal of Consulting and Clinical Psychology* 60 (3): 426–32.

Maruna, Shadd, and K. Roy. 2007. "Amputation or Reconstruction: Notes on 'Knifing Off' and Desistance from Crime." *Journal of Contemporary Criminal Justice* 23 (1): 104–24.

Millenky, Megan, Dan Bloom, Sara Muller-Ravett, and Joseph Broadus. 2012. *Staying on Course: Three-Year Results of the National Guard Youth Challenge Evaluation.* Washington, DC: MDRC.

Mischel, Walter, Yuichi Shoda, and Monica L. Rodriguez. 1989. "Delay of Gratification in Children." *Science* 244 (4907): 933–38.

Nagin, Daniel S., and Greg Pogarsky. 2004. "Time and Punishment: Delayed Consequences and Criminal Behavior." *Journal of Quantitative Criminology* 20 (4): 295–317.

Pearson, Frank S., Douglas S. Lipton, Charles M. Cleland, and Dorline S. Yee. 2002. "The Effects of Behavioral/Cognitive-Behavioral Programs on Recidivism." *Crime & Delinquency* 48 (3): 476–96.

Ponniah, K., and S. D. Hollon. 2008. "Empirically Supported Psychological Interventions for Social Phobia in Adults: A Qualitative Review of Randomized Controlled Trials." *Psychological Medicine* 38 (1): 3–14.

Raine, Adrian, Kenneth Dodge, Rolf Loeber, Lisa Gatzke-Kopp, Don Lynam, Chandra Reynolds, Magda Stouthamer-Loeber, and Jianghong Liu. 2006. "The Reactive-Proactive Aggression Questionnaire: Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys." *Aggressive Behavior* 32 (2): 159–71.

Republic of Liberia. 2012. *Liberia: Poverty Reduction Strategy Paper.* Washington, DC: International Monetary Fund Publication Services.

Robinson, T. E., and K. C. Berridge. 2000. "The Psychology and Neurobiology of Addiction: An Incentive-Sensitization View." *Addiction* 95 (2): S91–117.

Saini, M. 2009. "A Meta-Analysis of the Psychological Treatment of Anger: Developing Guidelines for Evidence-Based Practice." *Journal of the American Academy of Psychiatry and Law* 37 (4): 473–88.

Sapp, Stephen G., and Wendy J. Harrod. 1993. "Reliability and Validity of a Brief Version of Levenson's Locus of Control Scale." *Psychological Reports* 72: 539–50.

Schochet, Peter Z., John Burghardt, and Sheena McConnell. 2008. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *American Economic Review* 98 (5): 1864–86.

Spinella, Mohamad. 2007. "Normative Data and a Short Form of the Barratt Impulsiveness Scale." *International Journal of Neuroscience* 117 (3): 359–68.

Vigil, James Diego. 2003. "Urban Violence and Street Gangs." *Annual Review of Anthropology* 32: 225–42.

Westfall, Peter H., and S. Stanley Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* Hoboken, NJ: John Wiley & Sons.

Wilson, David B., Leana Allen Bouffard, and Doris L. Mackenzie. 2015. "A Quantitative Review of Structured, Group-Oriented, Cognitive-Behavioral Programs for Offenders." *Criminal Justice and Behavior* 32 (2): 172–204.

Wood, Elisabeth Jean. 2008. "The Social Processes of Civil War: The Wartime Transformation of Social Networks." *Annual Review of Political Science* 11: 539–61.

Copyright of American Economic Review is the property of American Economic Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.