<div align="center">

Reducing crime and violence: Experimental evidence from

cognitive behavioral therapy in Liberia

Christopher Blattman, Julian Jamison, and Margaret Sheridan

Online Appendix

</div>

A   Baseline sample

A.1   Full summary statistics and balance tests

Table A.1 expands the balance table in the main paper for the full set of baseline covariates available and used in the treatment effects regressions.[1] Column 1 reports the sample mean for each covariate, and Columns 2 to 7 report the coefficients and p values on treatment indicators from ordinary least squares (OLS) regressions of each baseline covariate on three treatment indicators (one for assignment to each treatment arm) controlling for block fixed effects. Column 8 reports the p value from a joint test of significance of the three coefficients. Finally, at the base of the table we report the p value from a test of joint significance of all covariates from an OLS regression of each treatment indicator on all covariates (including that treatment group and the control group alone).

Of 171 coefficients (57 covariates and 3 treatment arms), 14 (8.2%) have a $p < .05$, and 28 (16%) have a $p < .1$. Within treatment arms the covariates are not jointly significant, as seen from the joint test reported at the base of the table. Furthermore, 12 (21.1%) of the tests of joint significance have a $p < .1$.

Table A.2 repeats the same balance analysis for the 947 subjects interviewed at endline. Of 171 coefficients (57 covariates and 3 treatment arms), 13 (7.6%) have a $p < .05$, and 20 (11.6%) have a $p < .1$. Within treatment arms the covariates are not jointly significant, as seen from the joint test reported at the base of the table. Furthermore, 12 (21.1%) of the tests of joint significance have a $p < .1$.

Overall, therefore, there is minor imbalance. We control for all baseline covariates in all treatment effects regressions in the paper to account for this.

A.2   Neighborhoods and recruitment

Table A.3 describes each of the study neighborhoods where we recruited, along with population estimates. We report the estimates of the number of all adult males, as well as our low-end estimates of the number of target males in each neighborhoods—men 18 to 35 in the bottom decile of income.

---

[1]We maintained the Phase 1 baseline survey for all Phases for the sake of consistency and completeness.

Table A.1: Baseline statistics and balance test

Baseline covariate	Sample Mean (1)	Assigned therapy Coeff. (2)	Assigned therapy *p* value (3)	Assigned cash Coeff. (4)	Assigned cash *p* value (5)	Assigned both Coeff. (6)	Assigned both *p* value (7)	F-Test *p* value (8)
			Test of randomization balance (continued) (N=999)					
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	0.66	0.04	0.75	0.93
# of partners	0.53	0.05	0.45	0.12	0.18	-0.20	0.12	0.11
# of children<15 in household	2.21	-0.58	0.07	-0.51	0.19	0.64	0.29	0.32
Sees family often	2.37	0.12	0.00	0.24	0.01	-0.29	0.01	0.01
Muslim	0.10	0.02	0.64	0.00	0.96	0.01	0.87	0.18
Years of schooling	7.72	-0.20	0.68	0.03	0.95	-0.01	0.99	0.55
Currently in school	0.06	-0.03	0.08	-0.03	0.13	0.04	0.11	0.16
Literacy index (0-2)	1.23	0.14	0.12	0.12	0.30	-0.27	0.08	0.13
Math score (0-5)	2.79	-0.10	0.25	-0.03	0.85	-0.15	0.39	0.89
Health index (0-6)	4.87	-0.09	0.11	-0.19	0.17	0.31	0.15	0.28
Has any disabilities	0.08	0.04	0.29	0.00	1.00	-0.04	0.48	0.19
Depression index (0-17)	7.09	0.18	0.41	-0.01	0.97	-0.11	0.80	0.45
Distress index (0-21)	7.46	0.14	0.44	-0.01	0.98	-0.36	0.30	0.42
Relations to commanders index (0-4)	0.45	0.00	0.93	0.07	0.42	-0.06	0.55	0.72
Ex-combatant	0.38	0.06	0.09	0.08	0.08	-0.07	0.11	0.12
War experiences index (0-12)	5.86	0.34	0.25	0.15	0.54	-0.63	0.13	0.32
Weekly cash earnings (USD)	17.02	-1.89	0.03	-4.85	0.03	5.48	0.00	0.02
Summary index of income, z-score	0.00	-0.22	0.05	-0.12	0.48	0.26	0.21	0.07
Homeless in past two weeks	0.24	-0.01	0.80	0.00	0.93	-0.02	0.75	0.26
# of days slept hungry, last 7 days	1.26	0.25	0.10	0.28	0.05	-0.32	0.09	0.14
Savings stock (USD)	33.75	-10.08	0.26	-12.84	0.31	15.69	0.31	0.53
Can get loan of 50 USD	0.52	-0.03	0.59	-0.06	0.28	0.05	0.46	0.50
Can get loan of 300 USD	0.11	-0.03	0.27	-0.03	0.33	0.06	0.07	0.13
Hours in illicit activities	13.55	1.21	0.68	-0.86	0.67	0.06	0.99	0.14
Hours/week in agriculture	0.36	0.34	0.26	-0.10	0.35	0.13	0.84	0.01
Hours/week in low-skill wage labor	19.39	0.54	0.88	1.24	0.73	-0.43	0.90	0.94
Hours/week in low-skill business	11.53	0.16	0.92	-1.53	0.60	5.76	0.13	0.50
Hours/week in high-skill work	1.51	-0.05	0.91	0.94	0.03	0.11	0.85	0.01
Years of experience in agriculture	0.78	-0.21	0.29	-0.34	0.07	0.25	0.32	0.15
Years experience in non-agricultural business	2.96	-0.35	0.36	-0.80	0.05	0.97	0.08	0.04
Years experience in high-skill work	0.96	-0.29	0.13	-0.27	0.41	0.62	0.12	0.02
Sells drugs	0.20	0.01	0.69	0.00	0.92	0.00	0.93	0.92
Drinks alcohol	0.75	0.08	0.19	0.07	0.23	-0.07	0.23	0.31
Uses marijuana	0.59	0.12	0.02	0.08	0.02	-0.14	0.01	0.01
Uses marijuana daily	0.44	0.08	0.15	0.04	0.13	-0.09	0.22	0.34
Use hard drugs	0.26	-0.01	0.81	0.02	0.58	-0.01	0.81	0.83
Uses hard drugs daily	0.15	-0.04	0.21	0.02	0.52	0.01	0.90	0.37

Continued on following page.

Table A.1 (continued): Baseline statistics and balance test

	Sample	Test of randomization balance (N=999)						
		Assigned therapy		Assigned cash		Assigned both		F-Test
Baseline covariate	Mean	Coeff.	p value	Coeff.	p value	Coeff.	p value	p value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Committed theft/robbery in past 2 weeks	0.53	0.05	0.51	0.01	0.61	-0.02	0.62	0.77
Number of nonviolent stealing incidents	5.08	-0.36	0.58	-0.47	0.67	0.40	0.69	0.87
Number of felony stealing incidents	0.43	0.06	0.77	0.17	0.60	-0.17	0.67	0.86
Disputes and fights in past 2 weeks (0-9)	2.16	0.14	0.80	0.32	0.64	-0.66	0.25	0.69
Aggressive behaviors (mean of 19), z-score	0.00	0.05	0.66	0.13	0.22	-0.23	0.09	0.23
Conscientiousness index (0-24)	15.33	-0.05	0.70	-0.20	0.30	0.00	0.99	0.09
Neuroticism index (0-21)	12.08	-0.07	0.76	0.16	0.63	0.10	0.85	0.33
Grit index (0-21)	13.79	0.07	0.54	-0.07	0.82	0.00	0.99	0.20
Reward responsiveness index (0-24)	14.67	-0.16	0.37	-0.04	0.92	-0.21	0.70	0.92
Locus of control index (0-24)	14.45	-0.09	0.77	-0.43	0.15	0.45	0.29	0.00
Impulsiveness index (0-21)	9.35	0.35	0.39	0.16	0.65	-0.79	0.09	0.35
Self esteem index (0-24)	13.47	-0.08	0.78	-0.11	0.65	0.12	0.75	0.89
Patience in game play index (0-6)	4.12	0.05	0.62	-0.08	0.71	0.05	0.89	0.87
Time inconsistency in game play index (0-6)	3.27	-0.22	0.02	-0.05	0.62	0.13	0.33	0.01
Risk aversion index (0–3)	1.57	-0.01	0.90	-0.05	0.60	0.10	0.41	0.64
Self-reported patience (mean of 7), z-score	0.00	-0.08	0.62	-0.13	0.25	0.15	0.42	0.33
Declared Risk Appetite (mean of 6), z-score	0.00	0.01	0.94	-0.02	0.88	-0.10	0.65	0.94
Cognitive ability (z-score)	0.00	0.15	0.05	0.14	0.20	-0.29	0.01	0.04
Executive function (z-score)	0.00	0.07	0.18	0.10	0.45	-0.25	0.06	0.16
R-Squared		0.16		0.12		0.35		
p Value on F-statistics on all covariates		0.50		0.44		0.50		

Notes: 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 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.

Table A.2: Baseline statistics and balance test for endline respondents

| Baseline covariate | Sample Mean (1) | Test of randomization balance (continued) (N=947) | | | | | | |
| | | Assigned therapy | | Assigned cash | | Assigned both | | F-Test |
		Coeff. (2)	p value (3)	Coeff. (4)	p value (5)	Coeff. (6)	p value (7)	p value (8)
Age	25.35	-0.23	0.49	0.05	0.89	-0.02	0.96	0.17
Married or partnered	0.16	-0.02	0.75	-0.03	0.70	0.03	0.82	0.89
# of partners	0.53	0.06	0.45	0.11	0.22	-0.19	0.15	0.20
# of children<15 in household	2.24	-0.58	0.05	-0.48	0.19	0.65	0.27	0.27
Sees family often	2.37	0.09	0.08	0.21	0.01	-0.25	0.02	0.02
Muslim	0.09	0.02	0.56	0.01	0.60	0.00	0.95	0.75
Years of schooling	7.73	-0.30	0.51	-0.07	0.88	0.14	0.78	0.55
Currently in school	0.06	-0.03	0.09	-0.03	0.13	0.04	0.16	0.20
Literacy index (0-2)	1.22	0.08	0.36	0.08	0.47	-0.19	0.13	0.32
Math score (0-5)	2.79	-0.18	0.15	-0.09	0.66	-0.08	0.71	0.76
Health index (0-6)	4.85	-0.07	0.30	-0.19	0.21	0.28	0.25	0.43
Has any disabilities	0.07	0.04	0.29	0.00	0.97	-0.05	0.49	0.22
Depression index (0-17)	7.08	0.03	0.90	-0.07	0.82	0.06	0.90	0.88
Distress index (0-21)	7.44	-0.01	0.95	-0.09	0.79	-0.13	0.74	0.86
Relations to commanders index (0-4)	0.45	0.00	0.99	0.04	0.61	-0.04	0.74	0.88
Ex-combatant	0.38	0.07	0.10	0.07	0.13	-0.07	0.17	0.20
War experiences index (0-12)	5.87	0.32	0.36	-0.01	0.97	-0.52	0.24	0.41
Weekly cash earnings (USD)	16.90	-1.94	0.11	-4.33	0.03	4.70	0.00	0.01
Summary index of income, z-score	-0.01	-0.20	0.03	-0.11	0.41	0.25	0.14	0.06
Homeless in past two weeks	0.24	0.00	0.90	0.01	0.75	-0.03	0.61	0.22
# of days slept hungry, last 7 days	1.27	0.26	0.06	0.30	0.02	-0.37	0.04	0.07
Savings stock (USD)	32.46	-7.30	0.34	-7.26	0.47	10.28	0.42	0.62
Can get loan of 50 USD	0.51	-0.02	0.67	-0.05	0.32	0.05	0.44	0.50
Can get loan of 300 USD	0.10	-0.02	0.42	-0.02	0.47	0.05	0.26	0.44
Hours in illicit activities	13.22	0.61	0.80	-0.69	0.68	0.03	0.99	0.24
Hours/week in agriculture	0.37	0.35	0.33	-0.16	0.21	0.14	0.86	0.02
Hours/week in low-skill wage labor	19.34	0.46	0.91	0.78	0.84	-0.28	0.95	0.99
Hours/week in low-skill business	11.73	0.18	0.90	-1.56	0.60	5.54	0.15	0.53
Hours/week in high-skill work	1.46	-0.19	0.76	1.12	0.03	0.46	0.57	0.02
Years of experience in agriculture	0.74	-0.20	0.28	-0.30	0.12	0.27	0.25	0.31
Years experience in non-agricultural business	3.03	-0.41	0.24	-0.86	0.03	0.98	0.07	0.02
Years experience in high-skill work	0.93	-0.15	0.13	-0.04	0.87	0.44	0.11	0.03
Sells drugs	0.20	0.00	0.97	0.00	0.99	0.02	0.69	0.96
Drinks alcohol	0.76	0.06	0.28	0.07	0.30	-0.07	0.18	0.35
Uses marijuana	0.59	0.11	0.03	0.08	0.07	-0.13	0.03	0.04
Uses marijuana daily	0.44	0.06	0.21	0.04	0.18	-0.07	0.25	0.38
Use hard drugs	0.26	-0.02	0.55	0.01	0.88	0.01	0.86	0.85
Uses hard drugs daily	0.14	-0.04	0.27	0.01	0.66	0.01	0.92	0.43

Continued on following page.

iv

Table A.2 (continued): Baseline statistics and balance test for endline respondents

| Baseline covariate | Sample Mean (1) | Test of randomization balance (N=947) | | | | | | F-Test |
| | | Assigned therapy | | Assigned cash | | Assigned both | | |
		Coeff. (2)	p value (3)	Coeff. (4)	p value (5)	Coeff. (6)	p value (7)	p value (8)
Committed theft/robbery in past 2 weeks	0.54	0.04	0.59	0.02	0.46	-0.03	0.56	0.65
Number of nonviolent stealing incidents	5.07	-0.54	0.50	-0.50	0.66	0.50	0.60	0.80
Number of felony stealing incidents	0.45	0.04	0.85	0.17	0.60	-0.17	0.67	0.85
Disputes and fights in past 2 weeks (0-9)	2.10	0.08	0.88	0.55	0.47	-0.70	0.26	0.57
Aggressive behaviors (mean of 19), z-score	0.00	0.00	0.96	0.11	0.23	-0.17	0.20	0.27
Conscientiousness index (0-24)	15.34	-0.07	0.64	-0.26	0.32	0.07	0.85	0.03
Neuroticism index (0-21)	12.10	-0.08	0.65	0.14	0.64	0.16	0.73	0.35
Grit index (0-21)	13.78	-0.02	0.84	-0.18	0.61	0.13	0.70	0.79
Reward responsiveness index (0-24)	14.66	-0.20	0.30	0.04	0.91	-0.18	0.75	0.55
Locus of control index (0-24)	14.43	-0.11	0.71	-0.54	0.09	0.49	0.26	0.00
Impulsiveness index (0-21)	9.37	0.38	0.39	0.26	0.50	-0.84	0.12	0.40
Self esteem index (0-24)	13.47	-0.11	0.71	-0.11	0.62	0.16	0.69	0.89
Patience in game play index (0-6)	4.10	0.10	0.28	-0.08	0.69	-0.01	0.96	0.69
Time inconsistency in game play index (0-6)	3.27	-0.23	0.03	-0.03	0.79	0.10	0.53	0.00
Risk aversion index (0–3)	1.57	-0.02	0.87	-0.02	0.79	0.10	0.32	0.73
Self-reported patience (mean of 7), z-score	0.00	-0.05	0.75	-0.10	0.32	0.10	0.55	0.42
Declared Risk Appetite (mean of 6), z-score	0.00	0.03	0.88	0.02	0.90	-0.15	0.43	0.86
Cognitive ability (z-score)	0.00	0.13	0.14	0.16	0.18	-0.31	0.02	0.09
Executive function (z-score)	-0.02	0.07	0.21	0.11	0.40	-0.27	0.11	0.25
R-Squared		0.16		0.11		0.35		
p Value on F-statistics on all covariates		0.73		0.61		0.56		

Notes: 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 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.

Table A.3: Study sample and treatment assignment by block and phase

Phase	Start date (MM/YY)	Block (slum)	Short description	Estimated males		Recruited			% recruits assigned to:			
				All	Target	No.	% all	% target	Therapy	Cash	Therapy & cash	Neither
1	12/10	Red Light	Peri-urban, along the main road from Monrovia to the northeast of the country, residential but the site of one of the major markets in the city, mixed income neighborhood.	23 422	1 171	100	0.4%	8.5%	25.0%	25.0%	25.0%	25.0%
2	06/11	Red Light	Peri-urban, along the main road from Monrovia to the NE of the country, residential but the site of one of the major markets in the city, mixed income neighborhood.	36 434	1 822	219	0.6%	12.0%	26.9%	25.1%	24.2%	23.7%
	06/11	Central Monrovia	The area consists of Mamba Point and West Point, one of the busiest business areas in Monrovia.	32 345	1 617	179	0.6%	11.1%	31.8%	19.0%	31.8%	17.3%
3	03/12	Clara Town	Peri-urban, the north of Bushrod Island, the transit point to counties of the northwest of Liberia, notorious for petty crime.	28 704	1 435	240	0.8%	16.7%	28.6%	27.4%	22.9%	21.1%
	02/12	Logan Town	Peri-urban, the middle of Bushrod Island, next to Freeport of Monrovia, many garages and small shops and booths.	22 100	1 105	86	0.4%	7.8%	26.7%	29.1%	19.8%	24.4%
	02/12	New Kru Town	Peri-urban, the south of Bushrod Island, next to Central Monrovia, lots of car-loaders and wheelbarrowers.	23 921	1 196	175	0.7%	14.6%	26.3%	26.7%	23.8%	23.3%
All				166 926	8 346	999	0.6%	12.0%	27.7%	25.1%	24.9%	22.2%

Notes: Total male population estimates come from the authors' calculations based on data from Liberia Institute of Statistics and Geo-Information Services (LISGIS). To get an estimated number of target males we assume half are in the age range of 18-35 and take the bottom 10% decile as our targets. Red Light 1 includes Gorbachop, Woodcamp, Reservoir, Pipeline, Soul Clinic, and Sugar Hill. Red Light 2 includes Turtle Base, Chicken Poultry, Ma Kebbeh Gas Station area, Sugar Hill, Bassa Town, Goba Chop Community, Morris' Farm, Bernard Farm, Pipeline Road, Zayzay Community, Coca Cola Factory Community, Plank Field Community, Banana Bush Community, Soul Clinic Community, and Wood Camp. Central Monrovia includes Mamba Point and West Point areas. New Kru Town also includes part of Calwell. Logan town also includes part of Mamba Point and West Point that are not covered in Phase 2.

A.3 Tracking and attrition

We achieved tracking rates of roughly 93% over a year.[2] Given that this was such a transient population, we took special measures to minimize attrition.

Tracking to reduce attrition   At baseline we were clear about our desire to stay in touch. We took photos and signature samples, and collected as many as ten different ways to contact each respondent. We documented contact information for each respondent, including all the places they said they sometimes stay, plus contact information for the network of people around them who have a more stable location. Respondents were often on the run from the police or other people, and so their contacts might be uncomfortable speaking to enumerators and revealing the respondent's location. Thus, after the baseline survey, we asked respondents to use the enumerator's phone to call their most stable contact and introduce the enumerator and study and give permission.

At each endline, enumerators would typically start with the phone numbers of the various contacts or respondent and try to arrange an appointment. Contacts received no financial incentive. Failing that they would begin visiting the various locations listed. A slight majority of respondents were found within a few hours. In other cases, all leads were cold and more extensive sleuthing and asking around the neighborhood was required. If someone had traveled or moved far away, enumerators either waited until they returned or traveled across the country to find them in person.

On the upper tail, it could take three to four days of physical searching to find the hardest-to-locate people. Enumerators only stopped searching when all possible leads had been exhausted.

Response rates   Table A.4 lists survey response rates by treatment group and survey wave (pooling the 2- and 5-week surveys, and pooling the 11- and 13-month surveys). It also reports the p-value from a t-test of the difference between the response rate in each treatment group and the control group. None of the differences are statistically significant, and all are within about a percentage point of the control group response rate. The control group response rate is a tiny bit lower in the 12–13-month surveys and a tiny bit higher in the short run ones. But none of these differences control for covariates or even strata fixed effects, as in the next table.

Correlates of attrition and compliance   We analyze the correlates of attrition in Columns 1 and 2 of Table A.5, which reports an OLS regression of an indicator for attrition on selected baseline covariates.[3] There are not significant differences in attrition by treatment group, substantively or statistically. Those who attrit are slightly wealthier and have poorer mental health. In all, the treatment indicators and covariance are jointly significant at $p = 0.045$ so attrition is not ignorable. This is one reason we control for covariates in all treatment effects regressions.

---

[2]Rates of 80, 90 or even 95 percent are not uncommon in developing country field experiments and panel surveys. For example, the Indonesia Family Life Survey reached 94% of households and 91% of target individuals after four years. The Kenyan Life Panel Survey made contact with 84 percent of target respondents over a seven-year period. Similarly, in the US, researchers were able to reach 98% of the Perry Pre-school children at age 19 and 95% at age 27. One reason is that a small sample is easier to track intensively. Another reason is that enumerator wages are lower in Liberia in the U.S. and this means that intensive sleuthing and tracking is affordable.

[3]We do so to reduce collinearity and thus ease interpretation. Results with full covariates draw similar conclusions.

Table A.4: Survey response rates by wave and treatment status

	Control	Treatment Only	Cash Only	Treatment + Cash	All
2–3 weeks					
# found	384	484	427	433	1728
# unfound	33	48	49	40	170
Response rate	92.1%	91.0%	89.7%	91.5%	91.0%
p-value vs. control		0.65	0.36	0.83	
12–13 months					
# found	404	520	474	472	1870
# unfound	36	40	26	26	128
Response rate	91.8%	92.9%	94.8%	94.8%	93.6%
p-value vs. control		0.65	0.18	0.18	
All					
# found	788	1004	901	905	3598
# unfound	69	88	75	66	298
Response rate	91.9%	91.9%	92.3%	93.2%	92.4%
p-value vs. control		1.00	0.84	0.48	

Notes: Survey response rates are calculated as the difference between the total number of respondents at baseline and the number of respondents "unfound" at each endline, all divided by the number of respondents at baseline. Here, "unfound" refers to both respondents we could not locate and those we did locate but who choose to not participate in the survey.

A.4 Treatment compliance

Figure A.1 displays the distribution of class attendance for those assigned to therapy. NEPI did not collect attendance data during the first week (three sessions), so for simplicity we assume that all participants who attended at least one session after week one also attended the first three sessions.
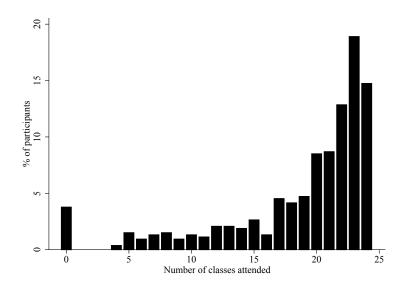
We use two definitions of compliance. Our first measure is defined as "attending at least 8 days of therapy", or about three of the eight weeks. Our second measure is defined as attending at least 80% of sessions (16 classes plus the 3 in the first week).

We analyze the correlates of compliance in Columns 3 through 8 of Table A.5. Being assigned to cash in addition to therapy did not affect the likeliness of attending therapy, which is to be expected since the cash grants were not known to participants until after therapy. The main correlates of compliance in the first three weeks are higher education, higher initial antisocial behaviors, and higher self-control skills. The main correlates of attending at least 80% of the sessions are higher education, better mental health, and patience in game play. Higher initial antisocial behaviors, and higher self-control skills are no longer so relevant.

Table A.5: Baseline correlates of survey attrition and treatment compliance for select covariates

	Dependent variable							
	Unfound		Cash Received		Attended >8d of therapy		Attended >19d of therapy	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Baseline covariate	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Assigned to therapy only	0.001	[.016]						
Assigned to cash only	-0.011	[.018]						
Assigned to therapy & cash	-0.003	[.016]	0.007	[.011]	0.012	[.026]	-0.002	[.042]
Age	0.001	[.001]	-0.003	[.002]*	0.000	[.003]	-0.005	[.005]
Married or living with partner	-0.015	[.020]	-0.005	[.021]	-0.055	[.037]	0.079	[.060]
# children under 15 in household	-0.003	[.002]	0.007	[.002]***	0.006	[.004]*	-0.005	[.007]
Years of schooling	0.001	[.003]	-0.001	[.002]	0.018	[.005]***	0.014	[.008]*
Cognitive skills, z-score	0.008	[.008]	-0.005	[.006]	-0.025	[.016]	0.025	[.025]
Health index (0-6)	0.001	[.005]	-0.001	[.003]	-0.008	[.008]	-0.008	[.015]
Symptoms of depression and distress, z-score	0.017	[.007]***	-0.001	[.005]	-0.012	[.011]	-0.056	[.023]**
War experience index (0-14)	-0.004	[.002]	-0.004	[.002]*	0.000	[.004]	-0.002	[.008]
Summary index of income measures, z-score	0.012	[.007]*	-0.008	[.008]	-0.016	[.015]	0.012	[.023]
Savings stock (USD)	0.000	[0000]	0.000	[0000]	0.000	[0000]	0.000	[0000]
Hours/week working in potentially illicit activities	0.000	[0000]	0.000	[0000]	-0.001	[.001]	-0.002	[.001]
Hours/week working (total)	0.000	[0000]	0.000	[0000]	0.000	[0000]	0.001	[.001]
Summary index of antisocial behaviors, z-score	-0.004	[.008]	0.005	[.007]	0.034	[.014]**	0.014	[.027]
Summary index of self-control skills, z-score	0.009	[.007]	0.001	[.005]	0.028	[.013]**	-0.001	[.024]
Patience in game play index (0-6)	-0.001	[.003]	0.003	[.004]	0.010	[.007]	0.021	[.011]*
N	999		499		529		529	
Mean of dependent variable	0.078		0.980		0.905		0.637	
R-squared	0.098		0.162		0.298		0.238	
P-value for test of joint significance	0.045		0.619		0.012		0.006	

Notes: Columns (1)–(2) pool all endline survey rounds and report the coefficients and standard errors from an OLS regression of an indicator for attrition on the baseline covariates. Columns (3)–(8) report the coefficients and standard errors from an OLS regression of an indicator for compliance on the same covariates, restricting the samples to people assigned to the respective treatment groups only. block fixed effects are included in all regressions but are omitted from this table. Robust standard errors are clustered at the individual level. *** p<0.01, ** p<0.05, * p<0.1

Figure A.1: Distribution of CBT Attendance

Notes: The figure reports the distribution of therapy attendance. No attendance data was collected during the first week, so we assume for simplicity that all participants who attended at least one session after week one also attended the first three sessions.

B Additional intervention details

B.1 Power calculations

After completing the pilot, we decided on a target sample of 1,000. This target was based on maximum program capacity and financial constraints. Based on the pilot, we estimated that the Minimum Detectible Effect for the full 1,000 (with a quarter for each treatment) would be a 0.12 standard deviation change in a standardized dependent variable for a two-tail hypothesis test with statistical significance of 0.05, statistical power of 0.80, an intra-cluster correlation of 0.25, and the proportion of individual variance explained by covariates as 0.10.

B.2 Randomization protocols

For the therapy and cash randomization, men in each block took turns drawing colored chips from an opaque fabric bag. In general, the bag was shaken and then the subject was instructed to turn away and to place one arm into the bag and to draw out a single chip. The color was confirmed and recorded.

In the cash instance, men were randomized in roughly equal sized blocks of about 50 people. Each man was invited into a private room to draw to ensure privacy and safety. This procedure was explained to the entire group, and all chips were placed into the bag in front of everyone. Then the bag was taken into a private room, and participants were called into the room individually. If they wished, they could inspect the bag to confirm that there were still chips of both colors inside. After everyone present had drawn, staff drew the remaining chips for the no-shows.

In the case of therapy, men were randomized each day, according to how many were recruited and surveyed in that neighborhood. This led to blocks ranging in size from 1 to 20, though the vast majority of blocks contained roughly 7 to 15 people. The draw was not as private as the cash draw, and men observed the outcomes of others drawing at the same time. Those who lost in the therapy randomization were offered a free meal along with the opportunity to discuss their situation with someone, and they were transported to a location of their choosing. A small percentage of the men were visibly upset and refused to engage at this point.

B.3 Therapy

NEPI's standard curriculum tended to be longer and broader than the two noncognitive skill and value changes that we study. For the purposes of this study, we worked with NEPI to streamline and focus the traditional STYL curriculum in two ways. First, we further grounded the approach in terms of CBT, emphasizing more practice over lectures. In general these modifications were quite modest, since the program already incorporated these techniques. Second, we asked NEPI to exclude modules not relevant to their theories of change: interpersonal skills; conflict resolution skills; dealing with war trauma and PTSD; career counseling; and community leadership.

To clarify and validate NEPI's curriculum, a Liberian qualitative researcher acted as a participant observer throughout one of the two Phase 1 pilot classes. Based on NEPI's training materials, our analysis of the theoretical grounding of the therapy, and this participant observation, we and NEPI developed a full program manual for the intervention.[4] The manual details the history and theory of the interventions, guidelines for recruitment of trainers and participants, training suggestions, the full curriculum, and guidelines for out-of–classroom engagement.

Curriculum

The curriculum has eleven main modules, which we present here with some examples of goals and activities:

1. *Transformation.* A tenet of CBT is that the therapist explicitly sets goals with participants and lays out the therapeutic strategy. This module introduces the concept of transformation, its significance, and the processes involved in transforming oneself.

   - The men are introduced to the techniques that will be used (role playing, lectures, storytelling, etc.), homework assignments, home visits, and the reasons for each.
   - The module also introduces ground rules for behavior, in terms of being respectful, practicing listening, waiting your turn, etc. The men do not necessarily have these skills, or haven't exercised them in some time, and learning to abide by these behavioral rules is an important part of the therapy.
   - Facilitators also begin to teach the songs, slogans, and call-and-response that will be used repeatedly throughout the course. These songs and slogans serve as important reminders of rules of behavior for the men to follow. They also can be used to bring order to a disorderly or inattentive group.
   - There are symbolic rituals to indicate a break in their lives. For example, the men write their "street names" and aliases on sheets of paper and they are burned together.

---

[4]Available at http://chrisblattman.com/documents/policy/2015.STYL.Program.Manual.pdf.

2. *Substance Abuse.* This module defines substance abuse and discusses its ill effects, as well as steps for moving past it. It explicitly encourages participants to reduce their consumption of drugs, alcohol, and tobacco. They are cautioned against cutting drugs entirely, to avoid withdrawal problems.

   - Men talk through and list reasons that they use drugs. The idea is to make them consciously aware of the reasons for their own behavior and risk factors in their lives. They also talk through the ill effects. Men talk through publicly about ways in which drugs have adversely impacted their own lives, sharing experiences.
   - Men role play situations where they could be pressured to use drugs and practice strategies for saying no.
   - An outside speaker comes to the classroom, often a former graduate of the therapy, to talk about their experiences with drugs and what it did to their lives, as well as what strategies they used to emerge. Men discuss strategies they can use in their own lives. They practice some of these as homework and come back to discuss their experiences with the class.

3. *Body Cleanliness.* The module explores the health, psychological, and social benefits of maintaining body cleanliness. Participants are encouraged to change behaviors that alienate them, and to present a public image (such as hair and dress) that promotes positive social interactions with community members.

   - Body uncleanliness is defined and highlighted as a problem mainly by getting men to discuss and volunteer their own opinions and experiences in a group.
   - The facilitators bring in a hair cutter, an electric shaver, and a set of nail clippers for men to clean up if they like.

4. *Garbage/Dirt Control.* An extension of the previous module, this module highlights the importance of cleanliness in participants' environments, and the ill effects of living in a dirty environment. It aims to help them maintain clean, healthy, and orderly living spaces.

   - Facilitators present the men with pictures of dirty and clean homes, businesses, and streets, and men point out different risks and unclean elements, and discuss the consequences.
   - Men identify ways they can improve cleanliness where they live (e.g. get a garbage can) and set and execute these plans as homework, to be followed up with home visits.

5. *Anger Management.* This module discusses the causes and effects of anger, the problems with acting out in ways they may later regret. It also provides participants with tools to manage their anger.

   - Men discuss the signs and indications of anger, in themselves and others, through discussion and role playing. Facilitators show pictures of angry faces and situations, and men interpret them. The aim is to make them cognizant of these signs.
   - Men discuss the causes of anger, and learn to link some of their actions to other people's anger.
   - Men discuss and role play the negative consequences of aggression and violence, or share experiences from their own life.

- Men practice nonaggressive responses to angry confrontations in class, such as learning to distract or calm oneself (walking away, doing other activities, starting discussions and de-escalating, or practicing breathing techniques). Men practice these techniques as homework.

6. *Self-Esteem.* This module emphasizes the need for participants to discover themselves in order to begin the path to recovery. This module links their behavioral changes to respect, pride, and confidence.

   - The facilitators try to link poor self-image directly to many of the behaviors they have discouraged in previous modules, both as a cause and consequence.
   - Men discuss ways they can build self esteem, make plans, and execute them as homework.
   - Facilitators work with men to identify worthwhile skills and characteristics they hold that are worthy of others' respect.
   - Men practice shopping for goods in a supermarket or shop as one of the first exposure activities. They work through successes and failures as a group and try again, sometimes with the help of a facilitator.

7. *Planning.* Reviews the steps and components necessary for planning and implementation. The goal of this module is to build participants' capacity to develop short- and long-term plans and understand the processes involved in executing these plans.

   - Planning skills are commonly taught in CBT programs as a method to build new skills. At its most basic, this involves helping the men break down larger plans into smaller steps and helping them work through ways to accomplish those steps, positively reinforcing successes and helping them process challenges and setbacks, often as a group. Men give examples and discuss them together. Another example: Small groups of men are tasked with organizing activities, such as a football match. The larger group listens to the different plans and critiques them.
   - As homework assignments, initially men are tasked with simple tasks (create a short term survival plan for feeding yourself or your family), and then more complex tasks (such as a business plan or home garden).
   - Men are also tasked with identifying a successful friend or family member and determining what steps led to their success. A motivational speaker (usually a past graduate) is also invited to talk about the steps involved in their success and their learnings and setbacks.

8. *Goal Setting.* The module outlines tools participants can use to develop goals, objectives, and indicators for measuring success in their own lives.

   - Participants are taught what short and long term goals are (through discussion and examples) and how to set reasonable short- and long-term goals (such as feeding their family, or starting a garden).
   - First participants practice setting goals and making plans, and then the larger group discusses and critiques them. Participants then set their own small, short term goals (e.g. changing a behavior, reconciling with a family member, or saving a certain amount this week) and execute these as homework, processing successes and failures as a group.

- Participants discuss the characteristics of good goals (e.g. achievable, measurable, time-bound) and revise goals and plans. They are given poor goals as a group and practice turning them into better goals. Another motivational speaker is used to discuss the role of goal setting in their own life.

9. *Money Business.* This module stresses the importance of engaging in positive spending habits and appropriately managing money. Impulsive spending habits are emphasized. Participants are taught to make plans and prioritize their needs and wants prior to spending their money.

   - Men engage in exercises to track their own recent spending to see where their money has gone. They discuss the use and misuse of their own money. As a group they discuss regrets and bad decisions and work through the negative consequences. These are illustrated dramatically through role-playing and skits, followed by discussion.
   - Later discussion, role playing and skits focus on techniques for resisting peer pressure and temptation. There is also testimony from a motivational speaker, usually a past graduate of the program.

10. *Money Saving.* The module introduces participants to various saving options and encourages them to reflect on the most suitable saving method for their lives. They practice interactions in informal and formal financial institutions.

    - Men discuss the reasons for and advantages of saving and it is explicitly linked to positive self image and esteem in the community. There is another motivational speaker.
    - Men learn techniques for saving safely at home without formal institutions. They learn to set and execute saving plans, using their goal setting and planning skills.
    - Homework assignments involve saving money they would have otherwise used on things they regret (identified in the previous module). Homework also involves trips to the bank and informal lenders. Prior to these assignments they meet and role play in groups, and their strategies are discussed and critiqued by the larger group. There is also a focus on appropriate presentation and image in these outings.

11. *Challenges and Setbacks.* The module explores potential challenges and setbacks they will face and has them practice positive coping mechanisms needed to effectively overcome them. Challenges and setbacks are framed as a test of one's maturity, potential, and abilities, and an opportunity for improvement.

A note on the approach

Note that in the United States, cognitive behavioral approaches to reducing violence are conscious of the fact that the values and behaviors it encourages could be maladaptive in some situations, since being violent can also protect people. As a result, these therapies teach people to judge when and where to use aggression.[5] NEPI, in designing the STYL therapy, did not consider the need for educating men on such contingent, adaptive behavior. Rather, their philosophy was that fighting back or retaliating in this context would lead to cycles of violence and an escalation of future risk,

---

[5]For instance, in a rough neighborhood, the optimal approach in terms of aggression is to retaliate when provoked, but to avoid starting a fight if not. Therapy aims to help people slow down their reactions and recognize when their automatic response (such as aggression) is and is not appropriate.

not a decrease. NEPI also emphasized how it was also important for the men who passed through STYL to demonstrate to the community that they were not aggressors or violent, to maintain the new image, and retaliation could be counter-productive there.

B.4 Cash grants

We contracted the international non-profit Global Communities (GC) to conduct the registration and cash distribution, as well as oversee NEPI's financial management and implementation schedule. We did so for several reasons:

1. To keep the therapy and the research teams distinct from cash distribution;

2. To coordinate registration and implementation of the two activities;

3. To relieve the research team of project and financial management of the interventions; and

4. To make the intervention as close as possible to a real-world, replicable intervention by other non-profit or state organizations.

For safety, GC developed a highly structured system of cash distribution. GC staff held cash in a car that moved around the neighborhood, to avoid theft. A lottery team with the men gave grant winners a voucher, and put them on a motorbike taxi that was then directed to the street corner where the car with the cash awaited. They were told to approach the car (which had an identifying mark such as a red bag on the dash), hand over their voucher, and receive their cash. The car would then move to a new corner, whose location would be relayed by mobile phone, and the process would repeat.

Anyone who was assigned to the cash treatment but was not present on the day of disbursal was still eligible for the grant. GC attempted to locate them for up to three weeks afterward, and generally succeeded.

C Formal theoretical model

Our model is rooted in previous models of occupational choice with self-employment (Fafchamps et al., 2014; Udry, 2010; Blattman et al., 2014), but adapted to have a criminal sector as in the broad class of models described by Draca and Machin (2015). We employed a similar model in Blattman and Annan (2015).

C.1 Setup

We model an individual's choice between legitimate business and illicit activities under different conditions—with and without time inconsistency, and with and without financial market imperfections—and assess the predictions for a number of common labor market and crime-reducing interventions: greater punishment, increasing productivity in legitimate business (e.g. through technology or skills improvement), cash or capital transfers, and interventions that shape preferences—either time preferences or personal preferences against illegal behavior.

We use L^b and L^c to denote time spent in legitimate activities (such as petty business) and illegitimate activities (such as crime). Legitimate business produces revenue according to production function $F(\theta, L_t^b, K_t)$, where θ is productivity or individual ability and K is accumulated capital used in business. A person's decision to participate in illegal activity is motivated by the potential gains and costs from such activity. Gains include the expected illegitimate payoff per hour spent in illegal activities, w. Costs include the possibility of apprehension and conviction, which occurs with probability, ρ, and implies a penalty, fL_{t-1}^c. Thus the penalty for criminal behavior is a linear function of hours spent in criminal activities in the previous period[6] The individual's total expected earnings from legitimate and illegitimate activities are $y_t \equiv F(\theta, L_t^b, K_t) + w_t L_t^c - \rho f L_{t-1}^c$. . In addition to investing in business, the individual can also invest or borrow through a riskless asset with constant returns $1 + r$. At each period t, the individual decides how much to invest for next period a_{t+1} and reaps interests ra_t from last period's investments.

Individuals have utility function $U(c, l, \sigma L^c)$, where c denotes consumption and l denotes time for leisure. We also allow for individuals to have direct disutility from engaging in crime, as measured by σL^c, where $\sigma > 0$ implies that implies that illicit work induces some internal penalty such as shame, though in principle it could also reflect social penalties such as a loss of esteem or exclusion from peers and other social networks. We make the standard assumption that $U_c' \geq 0$, $U_l' \geq 0$, $U_{\sigma L^c}' \leq 0$, $U_{cc}'' < 0$, $U_{ll}'' < 0$, $\partial^2 U / \partial L_c^2 \leq 0$ and $F_\theta' \geq 0$, $F_L' \geq 0$, $F_K' \geq 0$, $F_{\theta\theta}'' < 0$, $F_{LL}'' < 0$, $F_{KK}'' < 0$, and $F_{\theta L}'' \geq 0$, $F_{\theta K}'' \geq 0$, $F_{LK}'' \geq 0$.[7] We allow for the individual to have quasi-hyperbolic (β, δ) preferences.

We first consider the case without any uncertainty. The individual's problem is:

$$
\max_{c_t > 0, 0 \leq l_t \leq \bar{L}, L_t^b, L_t^c, K_{t+1}, a_{t+1}} \quad U(c_t, l_t, \sigma L_t^c) + \beta \sum_{i=1}^{\infty} \delta^i U(c_{t+i}, l_{t+i}, \sigma L_{t+i}^c)
$$

$$
s.t. \quad c_t + a_{t+1} + K_{t+1} \;=\; F(\theta, L_t^b, K_t) + w_t L_t^c - \rho f L_{t-1}^c + (1+r)a_t \quad \text{for each } t
$$

$$
a_0 \quad given
$$

where $L_t^b + L_t^c + l_t \equiv \bar{L}$.

---

[6] One reason for this modeling choice is because we want to explore the role that quasi-hyperbolic preferences play in the decision to commit crimes when the punishment is in the future not the present.

[7] For ease of analysis, we also assume that the marginal return to capital is infinity for the first unit of capital invested in business, and that as long as there is positive capital input, marginal product of labor for the first unit of labor will be infinity, i.e. $\lim_{K \downarrow 0} F_K'(\theta, L^b, K) = +\infty$ for all L^b and $\lim_{L^b \downarrow 0} F_K'(\theta, L^b, K) = +\infty$ as long as $K > 0$. This assumption guarantees that investments and hours in business will always be positive.

C.2 Occupational choice (and interventions) among time consistent individuals

Without credit constraints

Without time inconsistency ($\beta = 1$) or credit constraints, the set of optimality conditions are:

$$\frac{U_l'(t)}{U_c'(t)} = F_{L^b}'(t) \qquad if \; L_t^b > 0 \tag{1}$$

$$\frac{U_l'(t)}{U_c'(t)} - \sigma \frac{U_{\sigma L^m}'(t)}{U_c'(t)} = w_t - \frac{\rho f}{1+r} \qquad if \; L_t^m > 0 \tag{2}$$

$$1 + r = F_K'(t+1) \qquad if \; K_{t+1} > 0 \tag{3}$$

$$\frac{U_c'(t)}{U_c'(t+1)} = \delta(1+r) \tag{4}$$

$$c_t + a_{t+1} + K_{t+1} = F(\theta, L_t^b, K_t) + w_t L_t^c - \rho f L_{t-1}^c + (1+r)a_t \tag{5}$$

where for ease of notation, we use $U(t)$ to denote $U(c_t, l_t, \sigma L_t^c)$ and $F(t)$ to denote $F(\theta, L_t^b, K_t)$. Since we modeled crime punishment as a potential reduction in future wages, the risk neutral individual will view crime as an occupation with a discounted wage $w_t - \frac{\rho f}{1+r}$.

To find the marginal conditions for engaging in each sector, we first consider the case where illicit activity is not feasible. This would arise naturally if the probability of apprehension is high enough and punishment is heavy enough that $w \ll \frac{\rho f}{1+r}$. In this case the decision to engage in business depends on productivity θ, wealth level and the returns on other financial assets r. We use c^{ba}, L^{ba} and K^{ba} to denote consumption, labor and capital level in this scenario. Each period t, the individual chooses L_t^{ba} to satisfy $\frac{U_l'(c_t^{ba}, \bar{L} - L_t^{ba}, 0)}{U_c'(c_t^{ba}, \bar{L} - L_t^{ba}, 0)} = F_{L^b}'(\theta, L_t^{ba}, K_t^{ba})$ taking K_t^{ba} as given, and he chooses capital investment K_{t+1}^{ba} to satisfy $F_K'(\theta, L_{t+1}^{ba}, K_t^{ba}) = 1 + r$, taking expected L_{t+1}^{ba} as given.

Now, taking levels of c^{ba}, L^{ba} and K^{ba} as given, we then look at individuals' decision to engage in crime. Individuals will engage in illicit activities if and only if:

$$w_t - \frac{\rho f}{1+r} \geq \frac{U_l'(c_t^{ba}, \bar{L} - L_t^{ba}, 0)}{U_c'(c_t^{ba}, \bar{L} - L_t^{ba}, 0)} + \sigma \frac{-U_{\sigma L^m}'(c_t^{ba}, \bar{L} - L_t^{ba}, 0)}{U_c'(c_t^{ba}, \bar{L} - L_t^{ba}, 0)} \tag{6}$$

which says expected returns from crime are higher than the highest possible marginal rate of substitution between leisure and consumption the individual can achieve without engaging in crime. Since $-U_{\sigma L^m}'/U_c' > 0$, a rise in σ means more people will drop out of crime.

If condition (6) is satisfied and if $K_t > 0$, the individual then chooses L_t^b and L_t^c such that the marginal product of labor in business equals his expected marginal gains from crime, which also equals his marginal rate of substitution between leisure and consumption: i.e. conditions (1) and (2) will be satisfied. Notice L_t^c may not always be positive. The individual will not engage in crime if any or all three of the following happens: w_t is very low relative to the probability of apprehension ρ and punishment f; productivity in business θ is very high; the degree of aversion to crime σ is very high.

Capital investment and hours in business will satisfy condition (3). Notice that w, ρ and f will not affect returns to investment in business.

Interventions that increases the disutility of crime or the size or probability of punishment will reduce time devoted to in crime, but will have no effects on returns in business.[8] However, interventions that increase business productivity θ will not only induce more investment in business, but also reduce involvement in crime. In other words, $\frac{\partial L^c}{\partial \sigma} < 0$, $\frac{\partial L^b}{\partial \sigma}$ is ambiguous, $\frac{\partial L^c}{\partial \theta} < 0$ and $\frac{\partial L^b}{\partial \theta} > 0$. Finally, interventions that provide capital or liquid financial assets, such as a cash windfall, will not affect occupational choice at all, since the individual will already be working at his optimal level in both sectors. The windfall will simply be consumed and saved.

With credit constraints

In this section we consider the model with a simple credit constraint in the form of $a_t \geq 0$–individuals are unable to borrow in any period. We focus our attention on individuals whose initial a_0 is low enough that at some point in his life, the credit constraint is binding. Credit constraints will affect optimal conditions (2) and (3). The optimal condition for capital investment (3) becomes

$$F_K^{'}(\theta, L_t^b, K_t) = max\{1 + r, \frac{1}{\delta}\} \qquad if \ K_{t+1} > 0$$

and the optimal condition for hours in crime (2) becomes

$$\frac{U_l^{'}(t)}{U_c^{'}(t)} - \sigma \frac{U_{\sigma L^m}^{'}(t)}{U_c^{'}(t)} = w_t - \frac{\rho f}{max\{1 + r, \frac{1}{\delta}\}} \qquad if \ L_t^m > 0$$

Notice that $max\{1 + r, \frac{1}{\delta}\} \geq 1 + r$ and $w_t - \frac{\rho f}{max\{1+r,\frac{1}{\delta}\}} \geq w_t - \frac{\rho f}{1+r}$. For the impatient individuals whose $\frac{1}{\delta}>1+r$, their optimal level of capital investment will be lower than the baseline case because of the credit constraint. They are also have a higher expected returns from crime than in the baseline case, because the low level of business investment also forces them to put a higher discount rate on potential future punishment from crime.

Critical condition (6) becomes

$$w_t - \frac{\rho f}{max\{1 + r, \frac{1}{\delta}\}} \geq \frac{U_l^{'}(c_t^{ba}, \bar{L} - L_t^{ba}, 0)}{U_c^{'}(c_t^{ba}, \bar{L} - L_t^{ba}, 0)} + \sigma \frac{-U_{\sigma L^m}^{'}(c_t^{ba}, \bar{L} - L_t^{ba}, 0)}{U_c^{'}(c_t^{ba}, \bar{L} - L_t^{ba}, 0)}$$

Credit constraints induce more individuals who would otherwise not engage in crime to commit crime. For the impatient individuals, credit constraints increase their hours in crime and reduce their capital investments and hours in business activities.

Interventions that ease the credit constraint, including cash windfalls, will induce more investment in business and reduce involvement in crime. As in the baseline case, $\frac{\partial L^c}{\partial \sigma} < 0$, $\frac{\partial L^b}{\partial \sigma}$ is ambiguous, $\frac{\partial L^c}{\partial \theta} < 0$ and $\frac{\partial L^b}{\partial \theta} > 0$; however, the magnitude the effects of a change in σ or θ will be greater than in the baseline case; the magnitudes also increases with the degree of impatience: $\frac{|\partial L^c/\partial \sigma|}{d\delta} < 0$, $\frac{|\partial L^c/\partial \theta|}{d\delta} < 0$ and $\frac{|\partial L^b/\partial \theta|}{d\delta} < 0$ (notice that the lower the value of σ, the more impatient the individual).

---

[8]The level of investment in business may change depending on the shape of the utility and production functions, but the returns to investment will not change.

C.3 Occupational choice (and the effects of interventions) under time inconsistency

Without credit constraints

Time-inconsistent individuals ($\beta < 1$) will be more reckless in the present. Intuitively, the smaller is β, the more individuals want to enjoy higher consumption today at the expense of future consumption, which means they will borrow more, save less, invest less in business and/or involve more in criminal activities. However, as long as there is a perfect financial market, no one will change their business or criminal activities in order to consume more today–they will simply borrow more (or save less) today through the financial market.

In terms of optimal conditions, in the absence of any credit constraint, the only condition that changes is equation (4), which becomes

$$\frac{U_c'(c_t, l_t, \sigma L_t^c)}{U_c'(c_{t+1}^P, l_{t+1}^P, \sigma L_{t+1}^c)} = [\frac{\partial c_{t+1}}{\partial W_{t+1}}\beta\delta + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\delta] \cdot (1 + r)$$

where W_t denotes total wealth at time t, c_{t+1}^P denotes the individual's predicted future decision about c_{t+1} at time t. For the sophisticates $c_{t+1}^P = c_{t+1}$ while for the naifs $c_{t+1}^P > c_{t+1}$. Compared with the baseline case, the discount factor δ is replaced by the effective discount factor $\frac{\partial c_{t+1}}{\partial W_{t+1}}\beta\delta + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\delta$, a weighted average of the short-run and long-run discount factors $\beta\delta$ and δ where the weights are the next period marginal propensity to consume out of total wealth.

Notice that neither condition (2) nor condition (3) changes, as long as we have no credit constraints. Compared with the baseline, time inconsistency alone will not affect criminal activities or business investment. It would only change the level of savings or debts.

In this case, interventions that aim to correct time consistency will have no effects on either business investment or criminal activities, but will have an effect on consumption, savings and income.

With credit constraints

With credit constraints, in addition to equation (4), optimal conditions (2) and (3) will change as well. Let $\Delta = \frac{\partial c_{t+1}}{\partial W_{t+1}}\beta\delta + (1 - \frac{\partial c_{t+1}}{\partial W_{t+1}})\delta$ be the effective discount factor under (β, δ) preferences and $\tau = \frac{U_c'(c_t, l_t, \sigma L_t^c)}{U_c'(c_{t+1}^P, l_{t+1}^P, \sigma L_t^c)} \cdot \frac{1}{\Delta}$, where c_{t+1}^P denotes the individual's predicted future decision about c_{t+1} at time t. With credit constraints, the Euler equation (4) becomes

$$\tau \geq 1 + r \qquad with\ equality\ iff\ a_{t+1} > 0$$

and conditions (2) and (3) become

$$\frac{U_l'(t)}{U_c'(t)} - \sigma\frac{U_{\sigma L^m}'(t)}{U_c'(t)} = w_t - \frac{\rho f}{\tau} \qquad if\ L_t^m > 0$$

and

$$F_K'(\theta, L_t^b, K_t) = \tau \qquad if\ K_{t+1} > 0$$

In addition, critical condition (6) will change accordingly, with $1 + r$ replaced by τ.

Compared with the baseline case, $\tau > 1+r$ as long as an individual is credit constrained (i.e. has no savings). The level of τ will be higher for the sophisticates than for the naifs. However, regardless of their level of sophistication (i.e. the way individuals set their expectations for their future behavior), we know for sure that $\tau > \frac{1}{\delta}$, and the smaller β is (i.e. the more time inconsistent), the higher τ will be.

Compared to the time-consistent credit constrained case, fewer individuals will invest in business, more individuals will engage in crime, business investment levels will be lower, and hours in crime will be higher for everyone. The difference increases with the level of inconsistency (i.e. decreases with β).

Interventions that improve time consistency will shift people away from crime towards business. So will increasing the disutility of crime (though, as in the case without time inconsistency, while $\frac{\partial L^c}{\partial \sigma} < 0$, $\frac{\partial L^b}{\partial \sigma}$ is ambiguous). Increasing business productivity will have similar effects as before: $\frac{\partial L^c}{\partial \theta} < 0$ and $\frac{\partial L^b}{\partial \theta} > 0$. In all of these cases, however, the magnitudes the effects of a change in σ or θ will be greater than under time consistency, and the magnitudes also increase with the both degree of impatience and the degree of time inconsistency: $\frac{|\partial L^c/\partial \sigma|}{d\beta} < 0$, $\frac{|\partial L^c/\partial \theta|}{d\beta} < 0$, $\frac{|\partial L^b/\partial \theta|}{d\beta} < 0$, $\frac{|\partial L^c/\partial \sigma|}{d\delta} < 0$, $\frac{|\partial L^c/\partial \theta|}{d\delta} < 0$ and $\frac{|\partial L^b/\partial \theta|}{d\delta} < 0$. Notice that the lower the value of β, the more time inconsistent the individual is, and similarly, the lower the value of σ, the more impatient the individual is.

C.4   Introducing uncertainty and risk aversion

Three potential sources of risk are uncertainties in business productivity θ, wages from criminal activities w, and the potential punishment after apprehension f. We assume that decisions on business investment and hours in both sectors are made before risks are realized, and that θ, w and f follow independent stochastic processes.

With uncertainties in both the business and illicit sector, business investment and hours in both sectors depend on the variance of returns in both sectors and the level of initial wealth a_0. If both sectors are sufficiently risky, then those with high levels of wealth a_0 will turn away from both activities by reducing K, L^b and L^c and investing instead in other riskless assets. K, L^b and L^c will all be lower than the cases without risk. Those with low levels of initial wealth will not be able to live off savings alone, so they will have to invest more in either or both sectors, depending on the relative riskiness of the two sectors. As long as both sectors are similarly risky, K, L^b and L^c will all be higher; otherwise, if one of the sectors is less risky than the other, individuals will invest more time in that sector. $\frac{L^c}{L^b + L^c}$ will be lower than in the case without uncertainty if returns to crime are more volatile than business returns. One special case would be if individuals face a significantly positive chance of death after committing any crime. This is the equivalent of saying $f = +\infty$ with strictly positive chances. In this case hours in crime will be reduced to zero as long as the probability of apprehension is positive, $\rho > 0$.

With the presence of risk, inventions in θ will have greater effects, because an increase in θ now also makes business *relatively* less risky. A rise in σ will also have a bigger effect than without uncertainty, because risk aversion will reinforce the rise in aversion and further reduce hours in crime.

D Measurement

In this section, we discuss measurement decisions in more detail, including what was and was not specified in the 2012 National Science Foundation (NSF) proposal 1225697 that substitutes for the absence of a pre-analysis plan.[9] Section 4 of the proposal provides a numbered list of hypotheses and primary outcomes, and (roughly) how we planned to operationalize them, especially Sections 4.1 and 4.4. Section 5 expands on measurement approaches, both for these primary outcomes, as well as for control variables and other outcomes of interest. These are the key sections to examine now. Section 4 in particular is the basis for our organization of the current paper. That section and the introduction (Section 1) not only emphasize particular primary outcomes, but also the division into ultimate and intermediary outcomes.

We also report control group means and treatment effects on all of the survey questions that enter an index in the main tables. A note of caution: the standard errors have not been adjusted for multiple hypothesis testing, and so patterns across treatment effects within an index are suggestive only.

D.1 Antisocial behaviors

Table D.1 displays treatment effects for all components of our antisocial behaviors index.[10] These are purely illustrative, and we do not adjust standard errors for multiple comparisons.

Sections 1 through 4 of the NSF proposal make the primary, ultimate outcomes fairly clear: "poverty" and "violence", where Section 4.1.C defined "violence" as "crime, aggression, and political violence". As discussed in the main paper, only political violence was later dropped because none occurred before endline. We renamed this collection of outcomes "antisocial behaviors," for generality and clarity.

We are not aware of existing scales or measurement tools for Liberia, or even similar populations in sub-Saharan Africa or other low income countries. Thus, in general, our variables grew out of months of field work, qualitative interviews, and survey pre-testing by the authors and their research assistants, in order to understand common offenses and behaviors. Liberians speak a pidgin English and street youth have a slang of their own, and so even where we began with common scales (such as aggressive behaviors) the wording had to undergo extensive translation and testing to make sense. We also added new aggressive behaviors common to the study population and Liberian culture.

D.2 Economic performance

Table 3 of the main paper reports all measures of economic performance, and we do not replicate it here. The NSF proposal emphasized "poverty" as one of the two ultimate outcomes of interest, and in Section 4.1 expanded on this to discuss the expected impacts of the therapy on "economic decision-making and outcomes", including "levels of business investment and expenditures, savings, income and assets/consumption". The table in that proposal focused on business investments and our three

---

[9]See http://chrisblattman.com/documents/research/2012.01.13_STYL_NSF_proposal.pdf, 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.

[10]To save space, we only display 12–13 month results.

Table D.1: Program impacts on antisocial behaviors, 12–13 month survey only

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of antisocial behaviors, z-score	0.032	-0.083	[.093]	0.132	[.097]	-0.247	[.088]***
Usually sells drugs	0.135	-0.034	[.029]	0.035	[.030]	-0.059	[.029]**
# of thefts/robberies in past two weeks	1.839	0.073	[.395]	0.352	[.388]	-0.728	[.363]**
"Took something behind someone not for you" (stole)	0.275	-0.006	[.067]	0.106	[.072]	-0.063	[.069]
"Corrected someone's mistake" (stole unwatched items)	0.338	-0.067	[.077]	0.091	[.078]	-0.087	[.084]
"Scraped from others" (Cheating)	0.299	-0.037	[.071]	0.023	[.074]	-0.104	[.077]
Pick-pocketed someone	0.094	-0.039	[.039]	0.041	[.041]	-0.074	[.035]**
"Scammed someone" (Sold false goods or conned)	0.118	-0.005	[.038]	-0.044	[.039]	-0.093	[.036]**
"Black deed business" (Con artistry)	0.598	0.104	[.141]	0.016	[.128]	-0.196	[.137]
Mugged someone	0.086	0.075	[.077]	0.033	[.057]	-0.078	[.050]
Armed robbery	0.032	0.001	[.035]	0.021	[.036]	-0.026	[.026]
Disputes and fights in past two weeks, z-score	-0.060	-0.026	[.091]	0.100	[.090]	-0.100	[.077]
Small palava (dispute) with a neighbor	0.152	-0.005	[.062]	0.062	[.059]	0.047	[.065]
Small palava (dispute) with a leader	0.059	0.008	[.049]	0.005	[.041]	-0.008	[.035]
Small palava (dispute) with the police	0.152	0.033	[.064]	0.039	[.053]	-0.050	[.052]
Large fight with a neighbor	0.076	0.021	[.043]	0.023	[.033]	-0.036	[.032]
Large fight with leader	0.100	-0.027	[.047]	0.027	[.045]	-0.084	[.037]**
Large fight with police	0.027	0.021	[.027]	0.045	[.026]*	0.007	[.036]
Physical fight	0.115	-0.004	[.055]	0.087	[.065]	-0.079	[.045]*
Engaged in a fight with a weapon	0.083	-0.028	[.041]	0.023	[.042]	-0.048	[.037]
Fined for a fight	0.051	-0.081	[.062]	0.020	[.069]	-0.040	[.059]
Carries a weapon (typically knife)†	0.148	-0.059	[.031]*	0.043	[.035]	-0.066	[.033]**
Arrested in past two weeks	0.118	-0.006	[.024]	0.007	[.025]	-0.033	[.024]
Aggressive behaviors, z-score	0.188	-0.153	[.110]	-0.043	[.107]	-0.339	[.109]***
In the last 4 weeks, have you been quick to react against others?	0.611	-0.026	[.075]	0.047	[.078]	-0.117	[.078]
In the last 4 weeks, have you refused to take advice?	0.596	-0.113	[.081]	-0.042	[.082]	-0.183	[.082]**
Do you sometimes make hard jokes about people?	1.236	-0.013	[.093]	-0.115	[.094]	-0.156	[.092]*
In the last 4 weeks, have you intentionally destroyed property?	0.365	-0.026	[.073]	-0.057	[.072]	-0.110	[.074]
Do you sometimes cheat or scrape from people?	0.547	-0.047	[.084]	0.027	[.081]	-0.073	[.081]
In the last 4 weeks, have you ever had confusion with people about things?	0.606	-0.052	[.076]	0.016	[.079]	-0.123	[.082]

Continued on following page.

Table D.1 (continued): Program impacts on antisocial behaviors, 12–13 month survey only

	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
Outcome		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
In the last 4 weeks, did you let others see your frustration when you were frustrated?	0.555	-0.027	[.080]	0.015	[.080]	-0.085	[.082]
In the last 4 weeks, have you threatened other people?	0.360	-0.042	[.071]	-0.027	[.069]	-0.137	[.072]*
In the last 4 weeks, have you taken things from behind other people without asking them?	0.421	-0.047	[.075]	0.074	[.079]	-0.159	[.076]**
In the last 4 weeks, have you easily controlled your vexation when vexed? (-)	1.010	-0.051	[.090]	0.099	[.095]	-0.031	[.098]
Do you get vexed when you lose a game?	1.704	-0.163	[.094]*	-0.242	[.096]**	-0.251	[.095]***
Can you feel fine when you hit or yell at somebody?	0.828	-0.158	[.091]*	-0.051	[.092]	-0.269	[.092]***
If you are under attack can you hit that person to defend yourself?	1.867	-0.010	[.093]	-0.020	[.091]	-0.051	[.094]
When someone teases you, does that make you vexed?	1.493	-0.128	[.097]	-0.013	[.098]	-0.144	[.100]
Do you ever fight to show that you are the stronger person?	0.833	-0.104	[.094]	-0.011	[.094]	-0.153	[.098]
Do you ever damage things as a joke or for fun?	0.680	-0.162	[.084]*	0.010	[.083]	-0.250	[.084]***
Do you ever hurt the person you are playing football with for you to win?	0.882	-0.174	[.091]*	-0.177	[.090]*	-0.141	[.100]
Do you ever use force on somebody to do something for you?	0.567	0.041	[.086]	0.016	[.081]	-0.105	[.086]
Do you ever cuss somebody to do something for you?	0.596	0.054	[.084]	0.066	[.086]	-0.110	[.082]
Verbal/physical abuse of partner, z-score†	-0.071	0.142	[.100]	0.233	[.113]**	0.059	[.104]
Last month, did you accuse your woman for getting boyfriend?	0.388	0.115	[.083]	0.120	[.087]	-0.030	[.084]
Last month, did you ever tell your woman you will beat her?	0.381	0.084	[.078]	0.095	[.082]	0.027	[.079]
Last month, did you ever cuss your woman?	0.218	0.023	[.064]	0.123	[.075]	0.036	[.071]
Last month did you push, hit, slap or throw something at your wife or girlfriend?	0.152	0.076	[.056]	0.151	[.065]**	0.087	[.057]

Notes: The table reports 12–13-month intent to treat estimates of antisocial behavior outcomes. We calculate the impact of each treatment arm after 22-5 weeks and 12–13 months, controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1 †These variables were not collected during every phase/round, so their regressions have a smaller sample size.

measures of income (consumption, earnings, and asset stock). We look at all these measures of economic performance in a single family index, but would draw the same conclusions if we took a narrower definition of poverty and focused on the income measures alone, or even consumption alone.

D.3   Time preferences

Table D.2 displays control means and 12–13-month treatment effects for all subcomponents of our forward-looking time preferences index. The summary index consists of eight equally-weighted components: four measures for patience (δ) and four measures for time inconsistency (β). Components come from incentivized game play, hypothetical trade-offs over time, and survey measures.

The NSF proposal outlined these measures fairly specifically. Section 4.1.A specified our interest in the malleability of present versus future orientation and time inconsistency, and Section 4.4.A operationalized these measures as incentivized intertemporal choice games; hypothetical intertemporal choice games; and self-reported preferences. The main source of ambiguity was that the proposal referred to these measures variously as "discount rates", "present bias" or "forward-looking behavior".

In the end, the survey/incentivized games collected four types of measures, and each one yields a proxy of patience and time inconsistency

1. Incentivized trade-offs

   Following the survey, subjects were asked to play a set of "real money games" where they had to make a series of intertemporal choices between money at one point in time versus more money later in time, with some probability of a payout. The average payout was about \$3, roughly a day's wages.[11] The first choice was between money now and more money in two weeks; second between two weeks and four weeks; and finally one more question for each of these pairs of delays, but with the numbers modified depending on their first answer (i.e. if they chose to wait, then they were asked again but with a lower reward in the future). This bifurcating design allowed us to glean as much information as possible about their preferences with as few questions as possible, and we pretested the potential payouts to maximize the variance in responses.

   Based on game play, we assigned present and future patience scores for each respondent, ranging from 0 (less patient) to 3 (more patient).[12] We then used the sum of patience scores

---

[11]Subjects were told that one of the questions across the next few activities would be picked for payout, and their choice implemented, so that they should pay careful attention to their decisions. We told subjects that if one of the inter-temporal tasks was chosen for payout, and if their individual choice implicated a delayed reward, that we would come back and find them at the appointed time, in their own environment, to pay them.

Since we were typically returning in a few weeks to interview them again, and had interviewed them several times before, this was a reasonably credible commitment. Nonetheless, it could lead us to conflate patience with trust that the survey team would return. By the endline stage (their fifth survey with us), respondents knew us fairly well and knew that we were able to track them (and that we had paid them everything we had promised them in the past).

In fact, for logistical reasons, we also made one of the games a choice between a certain payout now and a lottery between a high and low payout (i.e. a risk preference question) and we selected this risk game for payout with very high probability, such that the intertemporal games were almost never paid out Although we did not technically lie at any point (since we did not mention the probabilities that each task would be paid out) this could be construed as minor deception. None of the respondents brought this up, even after having gone through the process five times.

[12]For example, if a respondent preferred 150 Liberian dollars (or LD, where 1 USD = 60 LD at the time) in a week over 50LD now, and 100 LD in a week over 50 LD now, they received a 3 for their present patience score. If they preferred 50 LD in two weeks over 150 LD in three weeks, and 50 LW in two weeks over 300 LD in three weeks, they received a 0 for their future patience score.

from the games to put people into 7 increasingly patient bins (0–6), and the difference of scores to put people into 7 increasingly time inconsistent bins.

2. Hypothetical trade-offs

   During the survey questionnaire, well before the incentivized games, we asked respondents to make the exact same series of tradeoffs as above, but in a purely hypothetical setting. We constructed the patience and time inconsistency proxies in exactly the same manner. Our aim was largely methodological, as we were interested in whether people responded differently when games were incentivized rather than hypothetical. This analysis—comparing the consistency and comparability of time preferences over different measures and over time—will be the subject of future methodological work, based on similar data we have collected across several countries and populations.[13] In the meantime, we merely use all available time preference measures in our summary index, in the interest of reporting all survey measures used from each family.

3. Hypothetical discount rate

   We also attempted to measure the discount rate in a second way (again, mainly for the methodological study mentioned above). As in Holt and Laury (2002), we asked respondents a series of hypothetical inter-temporal choices for larger amounts of money (on the order of US$10-30, about a week's wages). This was organized as two lists of 11 binary decisions, with a fixed amount right now versus a varying amount in two weeks (or two weeks versus four weeks for the second list). The delayed amount started as strictly less than the sooner amount (e.g. 1000 LD now or 900 in the future), then equal to, and then larger and larger until it was four times as big (1000 LD now or 4000 LD in future).

   We calculated discount rates based on each respondent's first switch from a present preference to a future preference.[14] Those who preferred 900 LD in the future over 1000 LD in the present received a discount rate of .9, while those who always preferred money earlier received a discount rate of 4. We then took the average of the inverse of the present (now versus 2 weeks) and future (in 2 weeks versus 4 weeks) discount rate as our measure of patience, and the difference between future and present as our measure of time inconsistency.

4. Self-reported survey questions

   We asked respondents six qualitative questions to gauge their self-reported levels of patience and time inconsistency.[15] For example, respondents were asked to place themselves on a ladder from 0 (least patient) to 5 (most patient) as one measure of self-reported patience, and how much they agree with statements such as "When I get money, I spend it quickly" as a proxy of time inconsistency. Specific questions are displayed in Table D.2.

By reporting all measures collected in the endline survey, three-quarters of our time preference measures are hypothetical rather than based on incentivized games. For robustness purposes, in Table D.2 we also report a summary index of the incentivized games only.

---

[13]In the meantime, we can see that the means similar (3.96 for the incentivized game versus 3.35 for the hypothetical), but this 15% difference is statistically significant at the 99% level.

[14]Enumerators continued down the list, and (oddly) a nontrivial fraction switched multiple times. We use the first switch only. Furthermore, about 17% of respondents preferred less money in the future as a commitment device, especially if they were expecting a large purchase coming soon.

[15]Dohmen et al. (2011) and Jamison and Karlan (2011) show that basic self-reported attitudes on risk and time preferences can be externally valid.

Table D.2: Program impacts on time preferences, 12–13 month survey only

		ITT regression (N= 947)					
	Control	Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of forward-looking time preferences	-0.149	0.149	[.102]	0.105	[.102]	0.209	[.105]**
Patience (δ) summary index, z-score	-0.240	0.170	[.103]*	0.145	[.096]	0.258	[.099]***
Incentivized trade-offs (0-6)	3.941	0.033	[.173]	0.266	[.166]	0.163	[.172]
Hypothetical trade-off (0-6)s	3.345	0.295	[.222]	0.376	[.221]*	0.331	[.222]
Hypothetical discount rate (.9 to 4)	2.238	-0.177	[.091]*	-0.121	[.091]	-0.251	[.091]***
Self-reported survey questions on patience, z-score	-0.363	0.091	[.093]	-0.032	[.092]	0.139	[.092]
Placing oneself on a 5-rung patience ladder (0-5)	3.757	0.003	[.109]	-0.048	[.107]	0.146	[.110]
"I consider myself a patient person." (0-3 scale)	2.267	0.091	[.063]	0.043	[.063]	0.129	[.065]**
"If I make good money, I save some for future problems." (0-3 scale)	2.108	0.041	[.072]	-0.078	[.072]	-0.035	[.073]
Time inconsistency (β) summary index, z-score	0.129	-0.072	[.083]	0.018	[.087]	-0.059	[.084]
Incentivized trade-offs (-3 - 3)	0.227	-0.024	[.071]	-0.010	[.074]	-0.038	[.073]
Hypothetical trade-offs (-3 - 3)	0.192	0.039	[.099]	-0.042	[.100]	-0.078	[.101]
Hypothetical discount rate (-3.1 - 3.1)	0.021	0.048	[.068]	0.060	[.072]	0.117	[.070]*
Self-reported (mean of 3), z-score	-0.473	0.185	[.100]*	-0.020	[.096]	0.136	[.103]
"If I get money, I spend it quickly." (0–3)	2.618	0.047	[.077]	0.020	[.075]	0.121	[.076]
"If I make good money, I spend a lot celebrating with friends." (0–3)	2.956	0.106	[.077]	-0.036	[.075]	0.070	[.079]
"I avoid going around friends who waste money." (0–3)	1.902	0.141	[.079]*	-0.025	[.079]	0.036	[.082]
Incentivized trade-offs (patience and time inconsistency), z-score	-0.047	0.029	[.081]	0.101	[.081]	0.084	[.084]

Notes: The table reports 12-13 month intent to treat estimates of outcomes that were not a priori specified as of primary interest. We calculate the impact of each treatment arm controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. The final variable in column 1 is the average of the patience and time inconsistency measures from our incentivized trade-offs. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

D.4 Self control skills

Table D.3 displays control means and 12–13-month treatment effects for all subcomponents and survey questions in our self-control index. Again, these treatment effects are for illustrative purposes only.

The NSF proposal highlighted "noncognitive skills of self control" as one of the three primary intermediate outcomes of interest in Section 4.1, though the proposal sometimes used "impulse control" or "self-discipline" synonymously. Sections 4.1.A and B gave examples such as "inhibition control, executive function, and perseverance". Section 4.4.A added that we will measure this using "standard psychological skills such as conscientiousness, locus of control, working memory, and inhibition." This ex ante description included executive function and locus of control, neither of which we presently include in family of self control measures in the paper. We decided to exclude these two measures after the NSF proposal was written but before final data collection. While this decision is not formally documented, the psychological and neurological principles are clear.

- While economists often refer to executive function as a "noncognitive" skill, one associated with self control, it is a technically cognitive ability in that psychologists and neuroscientists view it as a measure of mental performance established at a young age. Psychological and neuroscientific research suggests that executive function responds to childhood but not adult investments, and that investments result in very task-specific changes. Thus both theory and evidence suggests that this neurological capability should not be affected by CBT. Appendix D.7 below describes this research in more detail and illustrates the consequences of including or excluding executive function from the self control family.

- Locus of control or self-efficacy should never have been mentioned in the NSF proposal as linked to self control, and indeed the one mention of it was an aberration and an error. The concept of self control is intended to measure the degree to which you feel you are able to control your own emotions and behavior versus the degree to which you find you act impulsively or act as your emotions dictate without restraint. It is a skill. Locus of control, meanwhile, is a perception—a measure of the degree to which a person feels that events can be influenced by their own behavior versus luck or fate. It's possible that self control skills could affect a person's perceptions of self-efficacy. These two concepts are considered distinct, even unrelated, by psychologists. Psychologists view locus of control as a measure of self-regard and often combine it with measures such as neuroticism and self-esteem (Judge et al., 2002; Judge and Bono, 2001).

Instead, the survey included four psychological scales: impulsiveness, conscientiousness, GRIT, and reward responsiveness.

These existing scales typically have many more questions than we could use in the survey (or are commonly used in any assessment). These questions are typically organized into sub-scales to capture subcategories of behavior. We selected questions to use based mainly on whether they were easily understood and familiar to pre-test respondents, but we took care to ensure roughly equal proportions of questions from each sub-scale remained.

Because all personality questions were selected from questionnaires used in the United States, we first translated them into Liberian English by the enumerators, the authors and their research assistants then pre-tested the questions with young men from the same population as the youth in our study (but not members of the study sample).

To ensure that the questions continued to assess the original underlying constructs, we performed two checks. First, within the pre-test data we ensured that groups of questions were correlated or anti-correlated as one would expect given the underlying personality measure (e.g., impulsivity was negatively correlated with conscientiousness). Second, we performed a confirmatory factor analyses to ensure that within scales, questions were answered similarly.

D.5 Anti-criminal and anti-violent self-image/values

We are not aware of prior attempts to conceptualize or measure anti-criminal and anti-violent values and self-image. We developed three measures: self-reported values, prosocial behaviors, and appearance. The first two of these were prespecified in the NSF proposal.

Values and self-image were operationalized in Section 4.4.B of the NSF proposal, as the main "threat" to identifying the operation of the therapy through time preferences and self control skills. We recognized that the therapy could "change norms or beliefs about violence and its acceptability and risks and thereby reduce violence not through an effect on time preferences and self-control but through norms and the intrinsic utility or disutility of violent action." As a result, the proposal suggested that "we should observe a treatment effect of the [therapy] on self-reported norms towards violence, criminality and other antisocial behaviors, and possibly an increase in forms of collective actions, such as contributions to public goods or political participation." As we finalized the intervention, saw the pilot results, and designed the endline survey, we realized this was not a "threat" but interesting in itself, and simply another intermediary mechanism. In the paper, we chose to focus on the direct measure of preferences as opposed to the prosocial behavior (which were actions not preferences, and where by using the word "possibly" we indicated that we did not have strong priors).

As an afterthought, we also happened to measure the enumerator's impression of the respondent's appearance. We didn't conceive of these as measures of an actual skill or preference change, as they are choices or actions. Hence we treated them as an "other" outcomes in the last version of the paper. One of the referees argued persuasively that our measure of self-image should be changed ex post to include this measure of appearance. We agree and have made that change in the revised paper.

Self-reported values

The closest parallel to our measure of values is the measurement of social norms, where social psychologists ask respondents: (1) what the respondent thinks other people do (descriptive norms); (2) what the respondent thinks other people believe is appropriate (prescriptive norms); and, in some cases, (3) what the respondent him or herself believes is appropriate (an attitude) (Paluck, 2009). We used social norm surveys on behaviors such as bullying and conflict resolution as models for our approach, but had to develop our own original measures suitable to the context and treatment.

Prosocial behavior and appearance

From qualitative interviews (and prior surveys in the country) we also developed a number of locally-relevant prosocial behaviors and appearance measures

Table D.4 displays control means and 12–13-month treatment effects for survey questions in the identity and values index.

Table D.3: Program impacts on self control skills, 12–13 month survey only

		ITT regression (N=943)					
		Therapy only		Cash only		Both	
	Control	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
Outcome	mean						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of self-control skills	-0.070	0.159	[.090]*	-0.025	[.095]	0.244	[.095]**
Impulsiveness	0.082	-0.178	[.096]*	0.006	[.098]	-0.212	[.099]**
I buy things quick without thinking	1.246	-0.097	[.089]	0.012	[.091]	-0.082	[.092]
I can take action before thinking (-)	0.931	-0.025	[.091]	0.140	[.092]	-0.082	[.090]
I can just talk without thinking	0.744	-0.072	[.082]	0.021	[.086]	-0.208	[.082]**
I am not set or relax at lectures	1.153	-0.124	[.095]	-0.035	[.097]	-0.129	[.099]
I can catch hard time thinking	1.271	-0.074	[.093]	0.014	[.096]	-0.275	[.095]***
I believe in the present rather than the future	1.650	-0.150	[.105]	-0.190	[.106]*	-0.063	[.112]
I able to control myself (-)	0.596	-0.051	[.071]	0.066	[.078]	0.020	[.078]
I spend money on things and regret it later	1.848	-0.045	[.073]	0.017	[.074]	0.047	[.075]
Conscientiousness	0.018	-0.065	[.097]	-0.028	[.100]	0.044	[.097]
I am ready anytime	2.148	-0.267	[.067]***	-0.146	[.066]**	-0.215	[.067]***
I pay attention to things good good	2.305	-0.046	[.050]	-0.054	[.051]	-0.038	[.051]
I get everyday work done right away	2.044	0.006	[.062]	0.037	[.063]	0.108	[.065]*
I make plans and go by them	2.177	0.120	[.057]**	-0.011	[.057]	0.054	[.059]
I catch hard time to do my work (-)	1.704	-0.076	[.070]	-0.059	[.071]	-0.030	[.073]
I do unasked additional work after finishing my work.	1.468	-0.002	[.069]	0.075	[.072]	0.046	[.073]
I can't complete/finish things (-)	1.970	0.033	[.056]	0.026	[.060]	0.054	[.057]
I run away from work (-)	1.788	0.079	[.063]	0.066	[.062]	0.125	[.063]**
Perseverance/GRIT	-0.037	0.116	[.099]	0.057	[.099]	0.105	[.103]
I have overcome hard times to subdue an important challenge	2.020	-0.047	[.063]	-0.009	[.063]	0.060	[.068]
I can think big about my future	2.300	0.079	[.053]	0.094	[.053]*	0.114	[.054]**
Difficult conditions don't discourage me	1.980	0.026	[.065]	-0.090	[.068]	0.007	[.067]
My greatest prayer is to be successful in life	2.468	0.054	[.051]	0.100	[.052]*	0.026	[.055]
I'm trying hard to make it	2.305	0.009	[.053]	-0.008	[.053]	0.015	[.051]
I sometime make a plan but later on change my plan to a different one (-)	1.271	0.085	[.064]	-0.058	[.063]	-0.068	[.066]
I do not think too much about big things/success (-)	1.340	0.025	[.075]	0.086	[.075]	0.055	[.075]
Reward responsiveness	0.072	-0.165	[.102]	0.084	[.100]	-0.242	[.102]**
When I want something I can go to any corner to make sure I get it.	1.473	-0.149	[.075]**	-0.039	[.076]	-0.170	[.076]**
When I go after something, even the devil in hell can't stop me.	1.404	-0.072	[.078]	0.050	[.076]	-0.072	[.080]

Continued on following page.

Table D.3 (continued): Program impacts on self-control skills, 12–13 month survey only

Outcome	Control mean	ITT regression (N=943 responses from 943 subjects)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Most of the time, I will do things for no other reason than that I will enjoy them.	1.379	-0.138	[.059]**	-0.006	[.064]	-0.075	[.066]
When I am doing well at something, I will always like to be doing it.	2.059	0.061	[.053]	0.086	[.052]*	0.013	[.056]
When I get something I want, I can jump with happiness and it gives me plenty strength.	2.079	-0.027	[.056]	0.108	[.056]*	-0.096	[.060]
When I see a chance to get something I really like, I can jump with happiness on the spot.	1.966	-0.044	[.065]	0.013	[.065]	-0.129	[.064]**
When good things happen to me, it affects me strongly.	1.611	-0.023	[.069]	-0.082	[.071]	-0.057	[.071]
I can jump with happiness when I win a lucky ticket.	2.207	-0.053	[.057]	0.098	[.053]*	-0.070	[.059]

Notes: The table reports 12–13 month intent to treat estimates of self control outcomes. N=943 because 4 respondents did not answer all questions. We calculate the impact of each treatment arm controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table D.4: Program impacts on identity and values, 12–13 month survey only

	Control	ITT regression (N=947)					
		Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Summary index of identity and values, z-score	-0.041	0.036	[.094]	-0.074	[.092]	0.114	[.091]
Summary index of anticriminal and antiviolent values	0.070	-0.087	[.101]	-0.005	[.102]	-0.203	[.099]**
Attitudes toward use of violence (11 questions)	0.051	0.074	[.120]	0.029	[.122]	0.001	[.119]
If a stranger in the community robs one of your neighbors, is it okay for your neighbor to send people to abuse the stranger?	0.127	0.032	[.039]	-0.037	[.038]	0.003	[.040]
If a man owes you money but refuses to pay, would it be all right to take something from his home?	0.036	-0.009	[.026]	0.002	[.028]	-0.016	[.026]
If someone kills a known criminal in your community, it is ok if the police don't investigate?	0.122	0.022	[.035]	0.058	[.041]	-0.015	[.034]
If a rogue steals from a market, and the storekeepers chase him and beat him, should the storekeepers be punished? (-)	0.447	0.100	[.058]*	0.043	[.059]	0.052	[.058]
What if they mistakenly kill the rogue? Should the storekeepers be punished? (-)	0.189	0.011	[.045]	0.063	[.050]	0.038	[.048]
Suppose your friend's wife ran off with another man and stole his money and belongings. Would it be good for him to chase and beat that woman?	0.138	0.093	[.041]**	0.048	[.042]	0.020	[.042]
If a 'highman' is caught should his property be destroyed?	0.434	-0.043	[.057]	-0.036	[.059]	-0.002	[.059]
Suppose someone tries to steal your friend's girlfriend. Your friend is talking about getting the boys together to threaten the man. Would you join in?	0.107	-0.025	[.034]	0.005	[.035]	-0.035	[.034]
If a local leader is corrupt and taking money for himself, is it all right if he is beaten by the community?	0.163	-0.052	[.042]	-0.065	[.045]	-0.062	[.045]
If a wife challenges her husband in public, is it ok for her husband to beat her?	0.179	0.002	[.043]	-0.036	[.044]	-0.027	[.045]
Suppose a man rapes a girl in your community. He is arrested, but he bribes the police and goes free. If some people beat that police man, would you join in?	0.173	0.002	[.043]	-0.001	[.047]	0.037	[.046]
Attitudes toward criminality (12 questions)	0.044	-0.109	[.118]	-0.106	[.117]	-0.287	[.116]**
Imagine that a Chinese man cheats another man of his wage for carrying a load. Would it be ok for that man to take his friends to beat the Chinese man hard, and to take his money?	0.081	-0.049	[.026]*	-0.044	[.029]	-0.070	[.028]**

Continued on following page.

Table D.4 (continued): Program impacts on identity/values, 12–13 month survey only

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
If your best friend has some counterfeit money that need to be washed with mercury, will you join him in the search of the mercury to get lots of money?	0.168	0.009	[.042]	-0.002	[.045]	-0.052	[.042]
Is it wrong for someone in your community to make money by entering someone's house at night and taking their valuable things?	0.137	-0.070	[.038]*	-0.087	[.035]**	-0.042	[.041]
After being down on luck, a man finds a wallet in a taxi cab with lots of money and the driver's license. Should he take the money for himself before giving the wallet to the police or a radio station?	0.365	-0.063	[.054]	-0.021	[.058]	-0.098	[.057]*
Your family has no money to pay rent and will soon be homeless. You see $100 USD fall from a man's back pocket. Is it ok to take it and not tell him?	0.310	-0.022	[.053]	-0.019	[.053]	-0.048	[.054]
Is it okay for a man to hook up his house to an electricity cable from his family member to take free electricity behind his back, even though his family member will have to pay for it?	0.157	-0.026	[.044]	0.024	[.043]	-0.054	[.045]
A man's wife was in labour pain but he had no money. He came across a market woman's wallet filled with all her day's earnings. Is it okay for him to take the bag?	0.360	-0.009	[.055]	-0.058	[.054]	-0.113	[.055]**
If a man's $100 USD bill is hanging out of his back pocket and you have a clear change to jerk it with no one catching you, will you feel bad if you take it, even if you don't need it?	0.198	0.004	[.044]	0.001	[.044]	-0.067	[.043]
If a stranger left his room door open with all his valuables in there and no one is around, is it okay to correct his mistake?	0.107	-0.022	[.036]	-0.034	[.034]	-0.056	[.035]
Would you feel fine if your hustle was selling diazepam, bubble, 10-10, or any other drugs llike this?	0.137	-0.025	[.038]	-0.017	[.040]	-0.046	[.039]
If a friend left his room door open with his money inside and no one is around, is it okay to take his money?	0.091	-0.042	[.031]	-0.030	[.031]	-0.070	[.030]**
If you are working in a business, and you have access to your boss's money, would you feel fine taking that money for yourself, if you plan to return it later?	0.244	0.026	[.049]	0.005	[.051]	-0.046	[.051]

Continued on following page.

Table D.4 (continued): Program impacts on self-control skills, 12–13 month survey only

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Attitudes toward political violence (6 questions)	0.096	-0.143	[.118]	-0.002	[.123]	-0.225	[.119]*
Suppose the politician that you support does not win and there was no cheating. If that politician asks you to protest for him, would you do so?	0.129	-0.020	[.035]	-0.012	[.037]	-0.055	[.035]
Suppose the politician that you support does not win and there was no cheating. If that politician asks you to loot for him, would you do so?	0.079	-0.015	[.029]	0.009	[.032]	-0.042	[.029]
Suppose the politician that you support does not win and there was no cheating. If that politician asks you to loot for him for 200 LD, would you do so?	0.080	-0.011	[.031]	0.028	[.033]	-0.024	[.031]
Suppose the politician that you support does not win but there was cheating. If that politician asks you to protest for him, would you do so?	0.204	-0.075	[.042]*	-0.047	[.044]	-0.085	[.044]*
Suppose the politician that you support does not win but there was cheating. If that politician asks you to loot for him, would you do so?	0.151	-0.057	[.038]	-0.008	[.039]	-0.072	[.039]*
Suppose the politician that you support does not win but there was cheating. If that politician asks you to loot for him for 200 LD, would you do so?	0.130	-0.034	[.037]	0.024	[.037]	-0.056	[.037]
Quality of appearance (6 questions), z-score	0.016	-0.102	[.078]	-0.085	[.077]	-0.109	[.082]
Condition of clothes (0–2)	1.554	-0.063	[.050]	-0.069	[.051]	-0.037	[.051]
Condition of shoes (0–2)	1.569	-0.017	[.094]	0.005	[.093]	-0.058	[.097]
Cleanliness of breath (0–2)	0.713	-0.006	[.037]	0.004	[.035]	-0.036	[.038]
Cleanliness of face (0–2)	1.608	-0.077	[.047]	-0.077	[.046]*	-0.083	[.051]
Cleanliness of hair (0–2)	1.397	-0.028	[.056]	-0.001	[.055]	-0.026	[.056]
Cleanliness of fingernails (0–2)	1.520	-0.087	[.052]*	-0.090	[.051]*	-0.062	[.055]

Continued on following page.

Table D.4 (continued): Program impacts on identity/values, 12–13 month survey only

Outcome	Control mean	Therapy only ITT	Therapy only Std. Err.	Cash only ITT	Cash only Std. Err.	Both ITT	Both Std. Err.
		ITT regression (N=947)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prosocial behavior, z-score	0.018	0.041	[.088]	-0.075	[.085]	-0.017	[.090]
Number of active groups	0.692	0.062	[.085]	-0.157	[.083]*	-0.070	[.087]
Group leader index (4 questions)	0.049	-0.018	[.107]	-0.068	[.106]	-0.023	[.112]
Community leader now (0-1)	0.030	-0.003	[.019]	0.012	[.019]	0.003	[.019]
Big man in community (0-1)	0.108	0.002	[.034]	0.001	[.033]	0.017	[.035]
Group community leader (0-1)	0.193	-0.006	[.039]	-0.025	[.039]	-0.010	[.040]
Start group (0-1)	0.189	-0.012	[.040]	-0.050	[.041]	-0.036	[.040]
# of public good contributions (6 months)	2.500	3.352	[3.685]	0.484	[3.585]	4.092	[5.474]
Trust index (4 questions)	-0.079	-0.001	[.145]	0.075	[.142]	0.039	[.141]
Trusts relatives (0–3)	1.990	-0.171	[.158]	-0.091	[.158]	-0.100	[.154]
Trusts leaders (0–3)	0.796	0.001	[.140]	-0.005	[.136]	-0.126	[.140]
Trusts NGOs (0–3)	1.686	0.035	[.163]	0.051	[.160]	0.060	[.162]
Trusts IPA (0–3)	2.330	0.165	[.132]	0.280	[.125]**	0.255	[.132]*

Notes: The table reports 12–13 month intent to treat estimates of the identity and values index. We calculate the impact of each treatment arm controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

D.6 Additional intermediary outcomes of interest: Mental health, substance abuse, and social networks

Sections 1 to 4 of the NSF proposal made no mention of mental health, substance abuse, or social network change as major hypotheses or intermediary outcomes. Nonetheless, many of these were highlighted in Section 5 of the proposal as control or other outcomes of interest, as it is conceivable that they could be affected by the interventions, and that any change in these could also affect poverty or antisocial behavior.

Table D.5 displays the control group means and 12–13-month impacts of all the survey questions that comprise these three families. We describe the composition of each of these measures in Section 6.3 of the main paper.

D.7 Executive function

Table D.6 displays the control group means and 12–13-month impacts of all the components of executive function. We decided to measure executive function over time out of other research interests in measurement and consistency of executive function. We did not hypothesize a change in executive function, for reasons noted above, and so these results are not reported in the paper. This is the only surveyed outcome not reported in the paper's main tables. In the following section, below, we show the robustness of results to its inclusion.

In order to measure executive function, our behavioral protocol included three interactive activities

Table D.5: Impacts on mental health, substance abuse, and social networks

Outcome	Control mean	ITT regression (N=947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Positive self-regard/mental health, z-score	-0.050	0.032	[.091]	-0.034	[.091]	0.205	[.090]**
Positive self-regard	-0.054	0.029	[.093]	-0.035	[.092]	0.222	[.090]**
Neuroticism (8 questions)†	-0.019	0.044	[.097]	0.035	[.102]	-0.153	[.096]
Can feel sad one time (0−3)	1.951	0.044	[.049]	0.077	[.050]	-0.039	[.056]
Worry about things (0−3)	1.606	-0.049	[.067]	0.041	[.070]	-0.093	[.069]
Quick to feel threatened (0−3)	1.325	0.033	[.066]	-0.063	[.071]	-0.050	[.068]
Easily offended (0−3)	1.542	0.092	[.067]	0.051	[.072]	0.052	[.073]
Easily stressed (0−3)	1.340	-0.011	[.066]	-0.036	[.071]	-0.067	[.071]
Easily disturbed (0−3)	1.330	-0.013	[.064]	-0.056	[.068]	-0.076	[.065]
Feel relaxed most of time (-) (0−3)	1.355	-0.080	[.072]	-0.082	[.071]	-0.160	[.071]**
Nothing can bother me (-) (0−3)	1.034	0.092	[.062]	0.152	[.066]**	0.062	[.067]
Locus of control (8 questions)†	0.010	-0.032	[.101]	-0.111	[.098]	-0.022	[.106]
Your choices determine your future (0−3)	2.123	0.104	[.054]*	0.130	[.057]**	0.063	[.055]
You have small control over your life (-) (0−3)	1.163	-0.088	[.058]	-0.119	[.060]**	-0.105	[.061]*
Success in business is due to luck (-) (0−3)	1.300	-0.024	[.070]	-0.098	[.070]	0.060	[.078]
Trying hard can make your life better (0−3)	2.340	0.005	[.050]	-0.002	[.050]	-0.046	[.050]
Your can bring your plans into fruition (0−3)	2.123	0.076	[.048]	-0.039	[.045]	0.039	[.046]
Bad things in life are due to bad luck (-) (0−3)	1.616	-0.091	[.072]	-0.115	[.073]	-0.069	[.077]
People are homeless because of their own fault (0−3)	1.384	-0.011	[.074]	0.010	[.078]	-0.029	[.080]
Success comes from hard work (0−3)	2.379	-0.038	[.055]	0.003	[.059]	0.041	[.058]
Self esteem (8 questions)†	-0.071	0.078	[.098]	0.060	[.100]	0.190	[.101]*
Satisfied with yourself? (0−3)	1.872	0.119	[.069]*	0.017	[.072]	0.063	[.073]
Feel useless? (-) (0−3)	1.591	0.076	[.074]	-0.059	[.080]	0.137	[.077]*
Think everything will fail? (-) (0−3)	1.438	0.071	[.065]	0.082	[.066]	0.119	[.067]*
Don't get enough respect? (-) (0−3)	1.768	0.058	[.070]	0.071	[.073]	0.064	[.074]
Feel at least as good as most people (0−3)	1.985	-0.044	[.060]	0.045	[.059]	-0.049	[.066]
Feel like a good person but doing nothing (-) (0−3)	1.837	-0.052	[.067]	0.067	[.069]	-0.018	[.070]
Do business as well as most people (0−3)	2.202	-0.024	[.053]	-0.028	[.055]	0.011	[.053]
Feel shame (-) (0−3)	1.202	-0.010	[.073]	-0.047	[.078]	0.144	[.076]*
Summary index of subjective well being (3)	-0.020	0.057	[.072]	-0.009	[.072]	0.184	[.074]**
Absolute level today, 1-10 ladder	3.655	0.244	[.270]	0.024	[.271]	0.665	[.287]**
Overall level, (1-30)	13.198	0.504	[.711]	0.323	[.723]	1.912	[.746]**
Happiness (1−10)	4.000	0.274	[.285]	0.344	[.302]	0.837	[.295]***
Satisfaction (1−10)	4.315	-0.087	[.279]	0.053	[.294]	0.886	[.303]***
Health (1−10)	4.883	0.317	[.305]	-0.074	[.295]	0.188	[.307]
Relative to your community, (1-40+)	14.980	-0.026	[.821]	0.522	[.804]	2.642	[.859]***
Relative wealth (1−10)	3.035	0.081	[.224]	0.103	[.231]	0.606	[.248]**
Relative respect (1−10)	4.965	-0.034	[.312]	-0.023	[.316]	0.827	[.334]**
Relative power (1−10)	3.649	0.045	[.289]	0.334	[.269]	0.700	[.293]**
Relative access (1−10)	3.348	-0.138	[.249]	0.088	[.265]	0.489	[.256]*

Continued on following page.

Table D.5 (continued): mental health, substance abuse, and social networks, 12–13 month survey only

	Control	ITT regression (N=947)					
		Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Depression and distress	0.022	-0.031	[.093]	0.021	[.094]	-0.113	[.096]
Depression (6 questions)†	6.645	-0.072	[.302]	0.024	[.312]	-0.243	[.316]
Feels bad doing things you normally do (0–3)	1.256	-0.005	[.092]	0.053	[.095]	-0.028	[.100]
Hard time hearing people when thinking about bad things (0–3)	0.764	0.039	[.086]	-0.023	[.088]	-0.027	[.091]
Felt sad or down-hearted (0–3)	1.261	0.032	[.088]	0.084	[.087]	0.016	[.090]
Feels tired when even not doing anything (0–3)	1.409	-0.043	[.079]	-0.102	[.078]	-0.072	[.080]
Lost appetite from feeling bad (0–3)	1.108	0.007	[.080]	0.027	[.080]	-0.007	[.084]
Felt unimportant to everyone (0–3)	0.847	-0.084	[.090]	0.043	[.092]	-0.099	[.093]
Distress (12 questions)†	0.075	-0.036	[.095]	0.032	[.095]	-0.138	[.098]
Sits and thinks bad things	1.401	-0.108	[.095]	0.002	[.096]	-0.214	[.097]**
Has bad dreams	1.119	-0.219	[.090]**	-0.028	[.089]	-0.106	[.091]
Seems like bad things happening again	0.842	-0.108	[.088]	-0.023	[.089]	-0.098	[.092]
Sweat, feel warm, etc when thinking about bad things	1.000	0.000	[.091]	-0.033	[.086]	-0.178	[.091]*
Stay away from places	1.372	0.045	[.095]	-0.064	[.096]	0.037	[.098]
Talks to yourself	0.906	-0.052	[.092]	-0.043	[.090]	-0.219	[.093]**
Feels that your heart is spoiled	1.123	-0.060	[.086]	-0.020	[.084]	-0.158	[.089]*
Feels frustrated	0.881	-0.043	[.088]	0.133	[.090]	-0.047	[.091]
Thinks someone's spirit is hampering you	0.252	0.109	[.066]	0.120	[.065]*	-0.011	[.065]
Body is dry from worrying	0.940	-0.035	[.093]	0.014	[.095]	-0.029	[.098]
Heart does not feel satisfied	1.079	0.107	[.086]	0.065	[.086]	0.089	[.089]
Feels heart burning	0.765	0.121	[.086]	0.125	[.083]	-0.015	[.085]
Substance abuse (0-3 index)	1.091	-0.065	[.061]	0.063	[.062]	-0.057	[.060]
Usually drinks (0–1)	0.766	-0.063	[.042]	-0.047	[.042]	-0.045	[.043]
Usually uses marijuana (0–1)	0.490	-0.024	[.035]	0.018	[.035]	-0.042	[.036]
Usually takes hard drugs (0–1)	0.196	-0.005	[.031]	0.079	[.033]**	0.006	[.031]
Quality of social networks (z-score)	0.066	0.063	[.092]	-0.044	[.092]	0.139	[.095]
Peers (mean of 20 questions)	0.040	0.011	[.088]	-0.070	[.089]	0.017	[.090]
Friend in school (0–1)	0.802	0.003	[.039]	-0.060	[.043]	0.013	[.041]
Friend participates in community meetings (0–1)	0.624	0.005	[.047]	-0.049	[.049]	0.052	[.048]
Friend goes to church (0–1)	0.756	-0.013	[.043]	-0.064	[.044]	-0.048	[.045]
Friend works hard (0–1)	0.970	-0.077	[.026]***	-0.086	[.026]***	-0.083	[.027]***

Continued on following page.

Table D.5 (continued): mental health, substance abuse, and social networks, 12–13 month survey only

	Control	ITT regression (N=947)					
		Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friend has business or job (0−1)	0.660	-0.036	[.049]	-0.075	[.050]	-0.052	[.051]
Friend saves money regularly (0−1)	0.645	-0.051	[.048]	-0.017	[.049]	-0.043	[.050]
Friend gives good advice (0−1)	0.883	-0.027	[.033]	-0.070	[.035]**	-0.060	[.036]*
Friend likely to share (0−1)	0.827	0.032	[.037]	0.036	[.037]	0.033	[.038]
Friend cheers you up (0−1)	0.817	-0.045	[.041]	-0.002	[.040]	-0.053	[.041]
Fried trusted to guard valuables (0−1)	0.751	0.014	[.043]	-0.027	[.044]	-0.001	[.044]
Friend begs for money (-) (0−1)	0.772	-0.014	[.041]	0.003	[.042]	0.005	[.043]
Friend gets drunk regularly (-) (0−1)	0.787	0.002	[.042]	0.027	[.041]	0.032	[.042]
Friend uses drugs regularly (-) (0−1)	0.706	0.011	[.043]	-0.060	[.044]	0.015	[.043]
Friend pickpockets regularly (-) (0−1)	0.692	0.011	[.037]	-0.019	[.038]	0.027	[.038]
Friend burglarizes (-) (0−1)	0.782	0.050	[.039]	0.030	[.039]	0.056	[.040]
Friend is armed robber (-) (0−1)	0.909	0.037	[.028]	0.018	[.027]	0.064	[.026]**
Friend gambles (-) (0−1)	0.660	0.006	[.047]	-0.022	[.047]	0.038	[.047]
Friend is ex-combatant (-) (0−1)	0.548	-0.039	[.044]	-0.044	[.043]	-0.027	[.045]
Friend is former commander (-) (0−1)	0.721	0.015	[.043]	0.015	[.042]	0.064	[.042]
Friend has small fights with others (-) (0−1)	0.772	0.035	[.043]	-0.022	[.043]	0.007	[.044]
Friend has large fights with others (-) (0−1)	0.843	0.039	[.036]	0.030	[.036]	0.019	[.036]
Family (4 questions)	-0.019	0.124	[.099]	0.070	[.100]	0.129	[.097]
Sees family often (0−3)	1.904	0.129	[.094]	-0.021	[.096]	0.094	[.097]
Family concerned about you (0−3)	2.061	0.114	[.102]	0.090	[.103]	0.113	[.102]
Receives encouragement from family (0−3)	2.107	0.062	[.104]	0.048	[.102]	0.070	[.103]
Receives help from family when in trouble (0−3)	1.178	0.120	[.106]	0.132	[.104]	0.155	[.103]
Ex-commanders (4 questions)	0.176	0.004	[.076]	0.026	[.078]	-0.139	[.074]*
Friend is ex-military commander (0−1)	0.279	-0.015	[.043]	-0.015	[.042]	-0.064	[.042]
Close relations with ex-military commander (0−1)	0.208	0.018	[.042]	0.038	[.041]	-0.045	[.042]
Receives job from ex-military commander (0−1)	0.046	-0.006	[.021]	0.012	[.024]	-0.024	[.021]
Reports to commander now (0−1)	0.026	0.004	[.016]	-0.005	[.015]	-0.010	[.016]
"Big men" (5 questions)	0.120	0.001	[.155]	-0.130	[.152]	0.071	[.160]
Has patron for job (0−1)	0.238	-0.007	[.042]	0.011	[.045]	0.067	[.045]
Has patron for business needs (0−1)	0.099	0.008	[.030]	-0.042	[.029]	0.001	[.030]
Has patron for food (0−1)	0.347	0.041	[.047]	0.052	[.048]	0.051	[.050]
Has patron for school fees (0−1)	0.079	0.046	[.030]	0.028	[.030]	0.024	[.030]
Has patron for housing (0−1)	0.361	0.120	[.049]**	0.018	[.050]	0.070	[.052]

Notes: The table reports 12-13 month intent to treat estimates of mental health, substance abuse, and social networks. We calculate the impact of each treatment arm controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1 †These variables were not collected during every phase/round, so their regressions have a smaller sample size.

Table D.6: Program impacts on executive function, 12–13 month survey only

		ITT regression (N=947)					
	Control	Therapy only		Cash only		Both	
Outcome	mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Executive function (z-score)	0.098	-0.007	[.072]	-0.024	[.075]	-0.034	[.074]
Arrow time (z-score)	0.043	-0.034	[.045]	0.007	[.044]	-0.057	[.053]
Game 2	0.077	-0.013	[.010]	0.001	[.009]	-0.020	[.012]
Game 3	0.077	-0.015	[.033]	0.005	[.032]	-0.028	[.037]
Arrow error (z-score)	0.070	0.038	[.077]	-0.002	[.080]	0.045	[.078]
Game 2	0.258	-0.002	[.074]	-0.077	[.081]	0.004	[.076]
Game 3	0.179	0.066	[.078]	0.075	[.080]	0.073	[.076]
Maze (z-score)	0.060	-0.175	[.089]**	-0.116	[.088]	-0.162	[.091]*
First touch	0.010	-0.066	[.071]	-0.143	[.078]*	-0.100	[.073]
Backtrack	0.344	-0.189	[.088]**	-0.026	[.082]	-0.135	[.091]
Back digit	0.034	-0.053	[.071]	-0.060	[.074]	-0.108	[.075]

Notes: The table reports 12-13 month intent to treat estimates of executive function. We calculate the impact of each treatment arm controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1 †These variables were not collected during every phase/round, so their regressions have a smaller sample size.

drawn from economics and psychology.[16]

Planning behaviors We used a series of mazes to test planning behavior. Mazes were unknown to nearly all respondents. Subjects were shown an example maze on paper and then given 2, 2, and 3 minutes respectively to complete increasingly difficult mazes. Each had two entry points, one of which almost immediately led to a dead end. The main outcome of the mazes was the subject's ability to pause and plan their approach before completing the maze (i.e. did they plan their approach before choosing a starting point). As outcomes, we measure "time to first touch", or the amount of time spent planning prior to engaging in the maze; and number of mistakes (or "backtracks") in Maze 3, the hardest maze, which required the most planning and by which time participants had learned the concept of the maze. On average subjects took 18 seconds to plan for Maze 3 (SD = 23 seconds).

Behavioral inhibition and cognitive flexibility We developed the "arrows game", a modified directional Stroop task, a class of tasks that assess inhibitory control. Here subjects were shown a sequence of large black or white arrows that pointed either up or down and were first told to respond "up" or "down" to each arrow ("arrows baseline"). In the second version they were again shown the

---

[16]Across all behavioral tests administration was standardized. First, a clinical psychologist and economist trained enumerators in test administration. Next, in collaboration with experienced enumerators and research assistants, a comprehensive protocol was developed and used by all future enumerators. Enumerators were also instructed to answer clarifying questions and were taught the over-arching concept within each game so they could address questions/alleviate concerns without straying from the central concepts of the tests. This tight control over the testing situation allowed us to collect relatively sophisticated measures of cognitive function and behavioral responses to rewards in a constrained and otherwise under-resourced testing environment.

arrows but now were told to state the opposite direction; this constitutes producing the less common response while suppressing the more common response and is an assessment of inhibition ("arrows inhibition"). Finally, in a third version subjects were told to switch between two approaches: if the arrow was white they were to state the actual direction, but the opposite direction if the arrow was black. This is commonly called 'switching' and is an assessment of cognitive flexibility, the ability to move rapidly between two goals as the situation demands ("arrows switching"). For each version, the outcome data included total time to completion and the number of correct/incorrect responses out of 32 arrows. On average subjects made .33 errors (SD = 1.5) on arrows baseline, 2.4 errors (SD = 3.5) on arrows inhibition, and 3.9 errors (SD = 3.9) on arrows switching. Arrows took on average 25 seconds (SD = 17.7), 38 seconds (SD = 45.8), and 46 seconds (SD = 28.7) for baseline, inhibition, and switching separately.

Working memory Working memory is the ability to hold something in mind when it is no longer present in the environment and then manipulate it. The digit span task is an assessment of working memory. The digit span tasks involved the enumerator saying a random sequence of digits (1-9) out loud with a short pause between each digit, followed by the respondent repeating them back either in the same (forward-digits) or the reverse (backwards-digits) order. The enumerator began by giving two 2-digit numbers (one at a time) and recording the responses. If the subject correctly reported either of the numbers back, the enumerator would do the same with 3-digit numbers, and so on up to a maximum of 9 digits. As soon as the subject incorrectly reported both examples at a given level or span the enumerator moved on to the next activity (backwards-digits). The reverse digit span was done the same way, except that the subject was instructed to repeat the digits in the opposite order that the enumerator gave them (e.g., "three, zero, one") On average subjects were able to remember 5.5 digits forward (SD = 1.23) and 3.33 digits backwards (SD = 1.03). Each activity existed as two slight variants (e.g. changing the numbers in the gambles). These activities were alternated in the 2 versus 5-week endlines and the 12 versus 13-month endlines, so that participants were never asked identical questions too close together in time.

D.8 Distinguishing between different measures of "self-control"

Our summary indexes distinguish between self-control skills (assessed by various psychological scales), economic time preferences (using incentivized and hypothetical games), and (as an "other" outcome) executive function. Here we discuss the decision to separate these measures and what happens when we relax that assumption.

First, we treat the difference between time preferences and self control skills as an empirical question. As reported in Section 6.4, they are positively and significantly correlated but with a correlation of 0.33 it is unclear whether they are distinct or not. As we report in Table D.7, combining both into an equally-weighted index leads to large increases in the measure for both the therapy-only group (0.17 SD after 2–5 weeks, 0.18 SD after a year) and therapy and cash group (0.22 SD after 2–5 weeks, 0.26 SD after a year).

Second, we separate executive function from self control as well.

A main reason is that these abilities mature over the lifespan, and psychologists and neuroscientists have emphasized the importance of early-stage investments over late-stage investments because the neuroscientific principle of developmental plasticity, and data from randomizing young children into different early investments suggests that early, but not later investments shape cognitive function (Nelson, 2007).

This is not to say that they are not highly correlated or have common roots early in life. A large literature documents that in some extreme populations (e.g., individuals with substance abuse disorder, kids with ADHD) many of these indices of 'self control' co-vary. That is, kids with ADHD have deficits in performance on inhibition tasks (e.g. Barkley, 1997). These same children, by definition, behave impulsively and appear to be more sensation or risk seeking. Taken together, many have taken this covariance as evidence that these traits are interdependent. There is even a small neuro-imaging literature which suggests that these different forms of impulsivity are subserved by the same neural areas (Aron, 2007).

Nonetheless, there are many hints in the psychology and neuroscience literature that this is an oversimplification. For example, even within extreme populations, sensation seeking and impulsivity, measured similarly, may be differentially linked with behavior (Ersche et al., 2010). In typical developing children, successfully resisting temptation on delay of gratification tasks is not predicted by performance on inhibitory control tasks, but the strategies employed in attempting to resist temptation is (Eigsti et al., 2006).

In fact, the best test is to do what we have done here: randomly assign individuals to an intervention which shifts one of these indices and observe if they all move together. The fact that we see no improvement in executive function is consistent with the skills being different. In Table D.7 we test the combined measures formally, and we do not observe significant increases in a measure combining self control with executive function. Furthermore, their correlation is only 0.15, less than half of the correlation between self control and time preferences

E   Additional treatment effects analysis

E.1   Ignoring the ultimate/intermediary distinction

Table E.1 displays results for our main family indices if we ignore the ultimate and intermediary distinction and 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. Of course, if we were to adjust for eight comparisons within arms (e.g. if we were testing for specific hypotheses about each arm) statistical significant would be greater.

E.2   Robustness of treatment effects to alternate models

Our robustness tests focus on the five main summary outcomes. First, in Table E.2, we show robustness to alternative ways of constructing the indexes and pooling or averaging of endlines. Columns 2–4 report results from the main paper for comparison. Recall that in this main specification we averaged endline surveys (at 2 and 5 weeks, and 11 and 13 months), took an index of composite measures rather than individual survey questions and used equal weights. In columns 5–7, we do the same except use randomization inference to assess statistical significance. In columns 8–10, we pool our composite measures from both endline surveys and cluster our standard errors by individual. In columns 11–13, we do the same except weight each survey question equally. In columns 14–16, we use covariance-weighted indexes from Anderson (2008) and average both endlines.[17]

---

[17]For this index, each component is weighted by the inverse of the covariance matrix of all index components. Outcomes that are highly correlated with each other receive less weight while outcomes that are uncorrelated receive more weight as they represent new information. We cannot covariance weight the pooled endlines, since they are unbalanced in the sense that some outcome measures appear in only one endline while others appear in both.

Table D.7: Effects of combining various measures of self control

		ITT regression (N=947)								
		Therapy only			Cash only			Both		
Outcome	Control mean	ITT	Std. Err.	Adj. p-value	ITT	Std. Err.	Adj. p-value	ITT	Std. Err.	Adj. p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(7)
2–5 week impacts:										
Self control	-0.037	0.085	[.098]	0.833	-0.147	[.104]	0.595	0.037	[.096]	0.833
Time preferences	-0.202	0.179	[.098]*	0.419	0.071	[.099]	0.833	0.318	[.099]***	0.019
Executive function	-0.122	0.068	[.076]	0.952	0.032	[.082]	0.969	-0.011	[.090]	0.969
Self control + time preferences	-0.066	0.166	[.072]**	0.148	0.016	[.072]	0.882	0.221	[.073]***	0.029
Self control + executive function	-0.110	0.093	[.070]	0.671	-0.027	[.075]	0.913	0.009	[.080]	0.671
Self control + time preferences + executive function	-0.121	0.166	[.071]**	0.163	0.031	[.074]	0.676	0.163	[.076]**	0.209
12–13 month impacts:										
Self control	-0.070	0.159	[.090]*	0.631	-0.025	[.095]	0.992	0.244	[.095]**	0.154
Time preferences	-0.149	0.149	[.102]	0.826	0.105	[.102]	0.952	0.209	[.105]**	0.464
Executive function	0.098	-0.007	[.072]	0.999	-0.024	[.075]	0.999	-0.034	[.074]	0.999
Self control + time preferences	-0.212	0.183	[.088]**	0.343	0.060	[.088]	0.973	0.255	[.091]***	0.068
Self control + executive function	0.050	0.057	[.070]	0.974	-0.030	[.073]	0.991	0.068	[.072]	0.974
Self control + time preferences + executive function	-0.106	0.136	[.083]	0.667	0.032	[.084]	0.990	0.176	[.085]**	0.359

Notes: The table reports intent to treat estimates of the effect of each treatment arm controlling for baseline covariates and block fixed effects. Heteroskedastic robust standard errors are reported in brackets. Because there are two endline surveys per round per individual, each surveyed individual enters twice into the regressions, and standard errors are clustered by individual. *** p<0.01, ** p<0.05, * p<0.1

Table E.1: Program impacts on all eight family indices

		ITT regression (N= 947)								
		Therapy only			Cash only			Both		
Outcome	Control mean	ITT	Std. Err.	Adj. p-value	ITT	Std. Err.	Adj. p-value	ITT	Std. Err.	Adj. p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2–5 week impacts:										
Economic performance	-0.312	0.155	[.086]*	0.482	0.658	[.096]***	<0.001	0.580	[.096]***	<0.001
Antisocial behaviors	0.151	-0.249	[.088]***	0.007	-0.079	[.091]	0.886	-0.308	[.089]***	0.008
Time preferences	-0.202	0.179	[.098]*	0.482	0.071	[.099]	0.866	0.318	[.099]***	0.021
Self control	-0.037	0.085	[.098]	0.886	-0.147	[.104]	0.644	0.037	[.096]	0.886
Identity/values	-0.169	0.192	[.095]**	0.332	0.199	[.092]**	0.367	0.268	[.097]***	0.203
Self regard/mental health	-0.168	0.111	[.091]	0.203	0.137	[.093]	0.627	0.343	[.094]***	0.005
Substance abuse	1.378	-0.249	[.070]***	0.007	-0.060	[.072]	0.886	-0.197	[.071]***	0.075
Social network	-0.241	0.147	[.062]**	0.203	0.109	[.061]*	0.482	0.325	[.062]***	<0.001
12–13 month impacts:										
Economic performance	-0.031	0.065	[.104]	0.992	0.002	[.101]	0.997	0.069	[.098]	0.992
Antisocial behaviors	0.032	-0.083	[.093]	0.985	0.132	[.097]	0.920	-0.247	[.088]***	0.106
Time preferences	-0.149	0.149	[.102]	0.895	0.105	[.102]	0.985	0.209	[.105]**	0.546
Self control	-0.070	0.159	[.090]*	0.720	-0.025	[.095]	0.997	0.244	[.095]**	0.190
Identity/values	-0.021	0.013	[.089]	0.997	-0.101	[.089]	0.989	0.034	[.090]	0.952
Self regard/mental health	-0.050	0.032	[.091]	0.997	-0.034	[.091]	0.997	0.205	[.090]**	0.351
Substance abuse	1.091	-0.065	[.061]	0.983	0.063	[.062]	0.985	-0.057	[.060]	0.985
Social network	0.066	0.063	[.092]	0.992	-0.044	[.092]	0.995	0.139	[.095]	0.895

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 pre-defined 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. The overall family index is adjusted for 24 comparisons, Within each endline round, the component indexes are adjusted for 3 arms and the number of components within the family. Unadjusted p-values are indicated as follows: *** p<0.01, ** p<0.05, * p<0.1.

The conclusions from these three specifications are quantitatively similar to those from the main specification. Exceptions are as follows:

- The impact of cash and therapy on the covariance-weighted antisocial behaviors index is not significant after a year at conventional levels. This is because half of this index's weights come from domestic violence and number of arrests, two components that were unaffected by treatment. If we exclude domestic violence from the index and recalculate covariance weights, cash and therapy lead to a .26 standard deviation decline in antisocial behaviors after a year (column 19, significant at 99% level).

- Cash increases antisocial behaviors after a year in some specifications. In Column 15 we see that after a year the men who report cash only increased their antisocial behaviors by 0.17 standard deviations. In the other specifications, the coefficients are positive as well but smaller and not statistically significant. One possibility is that receiving a cash grant and failing, or having the money stolen, reinforces men's participation in crime. This is largely speculative, however.

Next, we check for robustness to alternative attrition scenarios by bounding treatment effects. We impute outcome values for unfound individuals at different points of the observed outcome distribution. The most extreme bound, from Manski (1990), imputes the minimum value for unfound treated members and the maximum for unfound controls. Following Karlan et al. (2015), we also calculate less extreme bounds by imputing relatively high values of the dependent variables for missing control group members, and relatively low values for missing treatment group members.[18] Specifically, we impute missing dependent variables for the treatment (control) group as the found treatment (control) mean minus (plus) 0.10, 0.25, or 1 SD of the found treatment (control) distribution. Note these imply large and systematic differences between missing treatment and control members—Columns 8 – 10 assume unfound control group member outcomes are roughly 2 SD greater than unfound treatment group member outcomes.

Table E.3 reports ITT estimates under these attrition scenarios. Our results are generally robust to these alternate specifications. When X = 0.25 SD, we still observe large and statistically significant changes in antisocial behaviors and our index of mechanisms after a few weeks and also a year. When X = 1 SD, our estimates of treatment effects lose significance but generally point in the correct direction. Meanwhile, the Manski bound brings us closer to having no treatment effects in the medium term term.

E.3 Both versus just one treatment

In this section, we compare the effects of receiving one treatment versus receiving both therapy and cash. Specifically, we test whether the coefficients on either therapy only or cash only in Section 6 are statistically different from the coefficients on therapy and cash. Table E.4 displays the mean difference between treatment effects and corresponding p-value for each of our three main outcome variables.

Our results indicate that cash and therapy compliment each other in reducing antisocial behaviors in the medium-run, while therapy compliments cash in the medium-run mechanisms.

---

[18]This assumes the dependent variable points in the positive direction. If treatment leads to a decrease in the outcome variable, as is the case for antisocial behaviors and antiviolent and anticriminal values, we impute in the opposite direction (i.e smaller values for control, larger values for treatment).

Table E.2: Robustness to alternative index construction and outcome measurement

Outcome, z-score	Round	Main specification			Endline averages, RI			Pooled mean effects of composite measures		
		Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Antisocial behaviors	2-5w	-0.249	-0.079	-0.308	-0.249	-0.079	-0.308	-0.201	-0.072	-0.260
		[.088]***	[.091]	[.089]***	[.082]***	[.082]	[.082]***	[.073]***	[.074]	[.072]***
	12-13m	-0.083	0.132	-0.247	-0.083	0.132	-0.247	-0.072	0.110	-0.219
		[.093]	[.097]	[.088]***	[.083]	[.083]	[.082]***	[.078]	[.080]	[.073]***
Economic performance	2-5w	0.155	0.658	0.580	0.155	0.658	0.580	0.156	0.647	0.530
		[.086]*	[.096]***	[.096]***	[.089]*	[.089]***	[.090]***	[.071]**	[.079]***	[.078]***
	12-13m	0.065	0.002	0.069	0.065	0.002	0.069	0.033	0.005	0.031
		[.104]	[.101]	[.098]	[.093]	[.094]	[.093]	[.075]	[.071]	[.071]
All mechanisms	2-5w	0.181	0.129	0.430	0.181	0.129	0.430	0.241	0.115	0.346
		[.094]*	[.092]	[.095]***	[.088]**	[.088]	[.088]***	[.063]***	[.062]*	[.064]***
	12-13m	0.107	-0.007	0.245	0.107	-0.007	0.245	0.112	-0.035	0.201
		[.093]	[.091]	[.092]***	[.087]	[.088]	[.087]***	[.080]	[.078]	[.079]**
N			947			947			1861	
Index of composite measures?			Y			Y			Y	
Endlines averaged?			Y			Y			N	
Randomization inference?			N			Y			N	
Covariance-weighted index?			N			N			N	

Continued on following page

Table E2 (continued): Robustness to alternative index construction and outcome measurement

Outcome, z-score	Round	Pooled mean effects of survey questions			Covariance-weighted average endlines			Covariance-weighted average endlines (no domestic violence)		
		Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both
	(1)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Antisocial behaviors	2-5w	-0.192 [.071]***	-0.073 [.071]	-0.270 [.070]***	-0.168 [.091]*	-0.076 [.091]	-0.277 [.085]***	-0.176 [.090]**	-0.094 [.090]	-0.292 [.083]***
	12-13m	-0.086 [.078]	0.041 [.078]	-0.257 [.072]***	-0.008 [.091]	0.173 [.097]*	-0.175 [.089]*	-0.085 [.099]	0.100 [.101]	-0.263 [.090]***
Economic performance	2-5w	0.156 [.071]**	0.647 [.079]***	0.530 [.078]***	0.129 [.083]	0.450 [.089]***	0.364 [.088]***			
	12-13m	0.033 [.075]	0.005 [.071]	0.031 [.071]	0.076 [.101]	0.004 [.092]	0.088 [.097]			
All mechanisms	2-5w	0.040 [.011]***	0.019 [.010]*	0.058 [.011]***	0.274 [.083]***	0.099 [.083]	0.393 [.085]***			
	12-13m	0.019 [.013]	-0.006 [.013]	0.034 [.013]**	0.152 [.089]*	-0.013 [.086]	0.255 [.087]***			
N			1870			947			947	
Index of composite measures?			N			Y			Y	
Endlines averaged?			N			Y			Y	
Randomization inference?			N			N			N	
Covariance-weighted index?			N			Y			Y	

Notes: The table reports the robustness of our results to alternate index construction and measurement outcome. In columns 2–4, we report results from our main specification, where we average composite measures and do not cluster standard errors. In columns 5-7, we do the same but use randomization inference to get our standard errors. . In columns 8–10, we pool our endline surveys, weight composite measures equally, and cluster standard errors by individual. . In columns 11–13, we pool our endline surveys, weight each survey question equally, and cluster standard errors by individual. In columns 14–16, we weight components using a covariance-weight from Anderson (2008) and average both endlines. In columns 17–19, we remove domestic violence from our antisocial behaviors index, weight survey questions using a covariance weight from Anderson (2008), and average both endlines,. *** p<0.01, ** p<0.05, * p<0.1

Table E.3: Robustness to alternative attrition scenarios

Outcome, z-score	Round	Impute missing dependent variable with mean = (-) X SD for missing control (treatment) individuals									"Worst case"		
		X = .10 SD			X = .25 SD			X = 1 SD			Manski bound		
		Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both	Therapy only	Cash only	Both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Antisocial behaviors	2-5w	-0.222 [.083]***	-0.063 [.084]	-0.292 [.083]***	-0.197 [.083]**	-0.038 [.084]	-0.266 [.084]***	-0.071 [.089]	0.087 [.091]	-0.137 [.090]	0.161 [.119]	0.318 [.123]***	0.109 [.121]
	12-13m	-0.076 [.088]	0.127 [.092]	-0.235 [.084]***	-0.060 [.088]	0.142 [.092]	-0.221 [.084]***	0.018 [.090]	0.219 [.094]**	-0.153 [.086]*	0.195 [.116]*	0.367 [.119]***	-0.005 [.121]
Economic performance	2-5w	0.142 [.080]*	0.614 [.089]***	0.538 [.089]***	0.120 [.081]	0.591 [.089]***	0.516 [.089]***	0.008 [.085]	0.477 [.095]***	0.402 [.094]***	-0.230 [.109]**	0.238 [.119]**	0.159 [.117]
	12-13m	0.041 [.100]	-0.008 [.095]	0.056 [.092]	0.025 [.100]	-0.024 [.095]	0.042 [.093]	-0.055 [.102]	-0.102 [.098]	-0.026 [.095]	-0.264 [.123]**	-0.301 [.119]**	-0.211 [.116]*
Mechanisms	2-5w	0.167 [.087]*	0.110 [.085]	0.414 [.088]***	0.144 [.087]	0.087 [.085]	0.390 [.088]***	0.030 [.091]	-0.029 [.090]	0.272 [.092]***	-0.238 [.118]**	-0.299 [.120]**	-0.013 [.120]
	12-13m	0.083 [.088]	-0.020 [.087]	0.235 [.088]***	0.068 [.089]	-0.035 [.087]	0.221 [.088]**	-0.011 [.091]	-0.113 [.089]	0.152 [.090]*	-0.164 [.105]	-0.252 [.103]**	0.026 [.105]
N		999			999			999			999		

Notes: The table reports intent to treat estimates of each treatment arm under alternative attrition scenarios. We impute missing dependent variables. In columns 2 – 10, we impute missing dependent variables for the treatment group as the found treatment mean minus a multiple of the standard deviation of the treatment distribution. Similarly, we impute missing dependent variables for the control group as the found control mean plus a multiple of the standard deviation of the control distribution. In columns 11 – 13 we apply Manski bounds, imputing the minimum value for unfound treated members and the maximum for unfound controls. Each regression controls for baseline covariates and neighborhood-phase fixed effects. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. . *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table E.4: Program impacts on therapy only versus therapy plus cash

	Dependent variable (N=947)					
Independent variable	Antisocial behaviors		Economic performance		Mechanisms	
	Mean diff.	p-value	Mean diff.	p-value	Mean diff.	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Both vs just therapy						
Short term survey	-0.06	0.44	0.43	0.00	0.25	0.01
Medium term survey	-0.16	0.04	0.00	0.96	0.14	0.13
Both vs just cash						
Short term survey	-0.23	0.00	-0.08	0.44	0.30	0.00
Medium term survey	-0.38	0.00	0.07	0.47	0.25	0.00

Notes: The table reports the mean difference between receiving both treatments and just one from intent to treat estimates, and corresponding p-values. We calculate the impact of each treatment arm controlling for baseline covariates and block fixed effects. We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1

E.4 2–5-week versus 12–13-month treatment effects

In discussing our results, we emphasize differences between outcomes 2–5 weeks after the intervention and outcomes 12–13 months after the intervention. In this section, we test whether the 2–5-week and 12–13-month impacts are the same. We pool our short-term results with our longer-term results and run the following OLS regression:

$$Y_{ij} = \beta_0 + \beta_1 ShortTerm_i + \beta_2 T_i + \beta_3 (ShortTerm_i \times T_i) + \mathbf{X}_i\lambda + \gamma_j + \omega_t + \varepsilon_{ij} \qquad (7)$$

where $ShortTerm$ is an indicator for outcomes measured in weeks 2 or 5, and T is an indicator for treatment group assignment. In our application, we have three treatment groups (therapy only, cash only, and therapy and cash), include baseline controls and block fixed effects, and cluster our standard errors at the individual level i. The size and direction of β_3 determine whether the treatment effects we observe after 2–5 weeks are the same as those observed after a year.

Table E.5 reports these estimates for our three main family indexes. For many outcomes, we cannot reject that β_3 is zero. In particular, the short- versus longer-term effects of both therapy and cash are not statistically distinguishable for antisocial behaviors and all mechanisms. However, there are two exceptions worth noting. First, while the cash-only group experienced the largest increase in the economic performance 2–5 weeks after the intervention, these effects diminished a year later. Second, while all three treatment groups saw decreases in antisocial behaviors in the short term, the effects of cash alone and therapy alone subsided 12–13 months later.

Table E.5: Testing the difference between 2–5-week and 12–13-month impacts of each treatment arm

	Antisocial behaviors		Economic performance		Mechanisms	
	β	S.E.	β	S.E.	β	S.E.
Independent variable	(1)	(2)	(3)	(4)	(5)	(6)
β_1						
Short term survey	0.106	[.067]	-0.200	[.064]***	-0.164	[.051]***
β_2						
Therapy only	-0.051	[.078]	0.046	[.075]	0.112	[.079]
Cash only	0.110	[.081]	0.005	[.073]	-0.049	[.075]
Both	-0.209	[.073]***	0.038	[.073]	0.216	[.077]***
β_3						
Short term X Therapy only	-0.165	[.082]**	0.092	[.086]	0.139	[.071]*
Short term X Cash only	-0.187	[.085]**	0.632	[.095]***	0.177	[.069]**
Short term X Both	-0.057	[.082]	0.489	[.090]***	0.133	[.072]*
Observations	3582		3618		3643	

Notes: We focus on pre-defined composite measures, typically defined by survey module. The overall summary indexes are the standardized mean of its composite outcomes, standardized. Heterosketastic robust standard errors are reported in brackets. We pool four endline surveys per individual (2 short-term and 2 medium-term) and cluster standard errors by individual. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table E.6: Impacts on crime incidence, in the last two weeks and annualized extrapolation

	Cash + therapy ITT, 12–13 month endline				Annualized impact	
	Control mean	Coeff.	Std. Err.	% change	Control mean	Cash + therapy
	(1)	(2)	(3)	(4)	(5)	(6)
# crimes, past two weeks	2.54	-0.994	[.438]**	-39%	66.1	-25.8
# times sold drugs, past two weeks	0.70	-0.266	[.188]	-38%	18.3	-6.9
# thefts/robberies, past two weeks	1.84	-0.728	[.363]**	-40%	47.8	-18.9
Selling/switching fake goods	0.27	-0.063	[.069]	-23%	7.1	-1.6
Stealing unwatched items	0.34	-0.087	[.084]	-26%	8.8	-2.3
Overcharging or cheating	0.30	-0.104	[.077]	-35%	7.8	-2.7
Burglary	0.09	-0.074	[.035]**	-79%	2.5	-1.9
Con artistry/scams	0.12	-0.093	[.036]**	-79%	3.1	-2.4
Pickpocketing	0.60	-0.196	[.137]	-33%	15.5	-5.1
Mugging	0.09	-0.078	[.050]	-91%	2.2	-2.0
Armed robbery	0.03	-0.026	[.026]	-82%	0.8	-0.7
Arrested in past two weeks	0.12	-0.033	[.024]	-28%	3.1	-0.8

Notes: Columns (1) to (4) report the same ITT regression as in Table 3, with robust standard errors in brackets. Columns (5) and (6) simply multiply the two week estimates by 26 weeks to generate an estimated annual impact per person.

*** p<0.01, ** p<0.05, * p<0.1

E.5 Crime: Disaggregated and annualized impacts

Table E.6 reports the incidence of specific crimes reported in the two weeks prior to the 12–13-month survey, breaking down the total number of crimes into the type of crime reported. For consistency, we shift from the incidence of drug selling reported in Table 3 to the frequency—the number of times men reported selling drugs in the past two weeks.

Control men committed 2.54 crimes in the previous two weeks, and this fell by almost one crime with therapy plus cash. All types of crime decreased by 20 to 100% with cash and therapy, but the statistically significant (and largest proportional) reductions are in burglary, muggings, and scams (e.g. the sale of non-existent goods, or down-payments for a hidden fortune). We do not adjust p-values for multiple hypothesis testing and so these comparisons across crimes should be taken with caution. In general the coefficients are negative and large in proportion to the control mean (20 to 100%) across all types of crime.

If this decline persisted for the year, it would translate to 26 fewer crimes per person each year. Given the $530 cost of the two interventions, this is roughly $21 per crime, ignoring any other benefits of the program.

E.6 Heterogeneity analysis on antisocial behaviors

Table E.7 reports impact heterogeneity from an OLS regression of the antisocial behaviors summary index on baseline level of either antisocial behaviors or self control and time preferences, treatment indicators, and interactions between treatment and baseline antisocial behaviors or an index of self control and time preferences, controlling for baseline covariates and block fixed effects. (Recall

that our measure of antisocial behaviors is a standardized index with mean zero. Therefore, the coefficient on the treatment indicator represents the treatment effect for an individual with mean level of antisocial behavior at baseline, while the coefficient on the interaction term is the additional effect for individuals whose baseline level of antisocial behaviors was 1 standard deviation higher than average.)

We did not prespecify any heterogeneity analysis with antisocial behaviors, and so these estimates must be taken with caution. But these were the only heterogeneity analysis we conducted.

Therapy decreased the incidence of antisocial behaviors for the average participant, but men exhibiting more antisocial behavior at baseline saw larger declines. For example, men with average levels of antisocial behaviors at baseline who were assigned to both therapy and cash experienced a 0.25 standard deviation decline in their level of antisocial behaviors 12–13 months later, but men whose initial level of antisocial behaviors was a standard deviation higher than average experienced about double the decline. Our results also indicate that after a year, men with high levels of initial antisocial behavior who received a cash grant actually increase their antisocial acts. This is especially interesting given that the effects of cash on occupational choice and income disappeared after a year. One possibility is that this increase in antisocial behavior is a reaction to the failed attempt at legitimate livelihoods, but these results are more speculative than anything else.

Our results also indicate that therapy and cash decreased the incidence of antisocial behaviors by 0.25 SD for participants with average self control and time preferences, but the effects were smaller for men who were more patient at baseline. These conclusions remain when we adjust for two comparisons within the "both" treatment arm.[19]

E.7 Program impacts on occupational choice

To measure changes in occupational choice, we asked respondents at each endline whether they had engaged in 22 occupations, from farming to petty business, trades, and formal jobs. For each occupation, we collected self-reported earnings and hours in both the last week and the week prior. We use these to calculate the total earnings and hours variables. With two endline surveys, we have four weeks of employment data per person in both the 2–5-week and 12–13-month surveys.

We can also calculate hours by occupations each week, aggregating our 22 occupations into 5 mutually exclusive categories:

1. Non-agricultural high-school work, which includes trading and office work

2. Non-agricultural low-skill business, which includes selling from a shop, selling at a table, buying and selling, engaging in petty trade, and conducting small business

3. Non-agricultural low-skill wage labor, which includes contract work, carloading, car-washing, peim-peim riding, carrying loads, guarding, housecleaning, and construction

4. Agricultural work, which includes farming and fishing,

5. Illicit work, which includes selling drugs, stealing, gambling, gold rubber, and scavenging.

---

[19]The adjusted p-values for the interaction terms on "assigned to both" are 0.011 in the short term and 0.014 in the long term for antisocial behaviors, and 0.177 in the short term and 0.037 in the long term for self control and time preferences.

1

Table E.7: Impact heterogeneity based on initial levels of antisocial behavior and self control/time preferences

| Baseline covariate | Round | Coefficient on: | Therapy only | | Cash only | | Assigned to both | |
| | | | Coeff. | Std. Err. | Coeff. | Std. Err. | Coeff. | Std. Err. |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Antisocial behaviors	2–5w	Treatment indicator	-0.247	[.087]***	-0.075	[.091]	-0.313	[.088]***
		Interaction term	-0.198	[.101]*	-0.047	[.103]	-0.253	[.091]***
	12–13m	Treatment indicator	-0.080	[.093]	0.138	[.097]	-0.248	[.088]***
		Interaction term	-0.042	[.107]	0.185	[.112]*	-0.230	[.085]***
Self control and time preferences	2–5w	Treatment indicator	-0.247	[.088]***	-0.078	[.091]	-0.310	[.089]***
		Interaction term	0.016	[.089]	0.023	[.093]	0.116	[.086]
	12–13m	Treatment indicator	-0.080	[.093]	0.127	[.096]	-0.250	[.088]***
		Interaction term	0.044	[.087]	-0.148	[.101]	0.168	[.080]**

Notes: We regress our family index of antisocial behaviors on baseline level of antisocial behaviors or self control/time preferences, treatment indicators, and interactions between treatment and baseline antisocial behaviors, controlling for baseline covariates and block fixed effects. Robust standard errors in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table E.8 reports ITT estimates on the average of the two weeks of data. While we generally observe no changes in overall average hours worked per week (the one exception is those assigned to cash only work approximately 15% more hours per week in the short-term), treatment effects how participants allocate their time. In the short-term, all three treatments cause participants to shift from illicit work to non-agricultural low-skill business. Those assigned to both therapy and cash experience the largest decline in illicit work. Time spent in illicit work falls 38% 2–5 weeks after implementation relative to the control group, and is 17% less than the control group one year later (although the latter is not statistically significant). Although the cash-only group more than doubles its weekly hours spent in non-agricultural low-skill business in the short-term, these effects phase out 12–13 months later.

E.8   Program impacts on baseline data

There is the threat that post-randomization outcomes could be due to baseline imbalance. Table E.9 investigates this by displaying program impacts on families of baseline variables aggregated to be as similar to our endline outcomes. The time preferences, self control, mental health, and substance abuse indices are the same as at endline. The antisocial behavior index is missing (i) carries a weapon on body, (ii) arrested in past 2 weeks, and (iii) verbal/physical abuse of partner because only collected data on these outcomes at endline. The economic performance index is missing the investment and non-durable consumption components. We collected none of the components of the identity and social network indices at baseline, so they are excluded from the table.

If baseline imbalance is driving our results, we should see treatment effects on baseline data. However, Table E.9 shows this is not the case. No treatment effect is significant at the 95% level. Only four comparisons out of 36 (11.1%) have $p < 0.10$, and none of these are in the therapy plus cash

Table E.8: Program impacts on occupational choice

			ITT regression (N=947)					
			Therapy only		Cash only		Both	
Outcome	Round	Control mean	ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average hours/week of work, past 2 weeks	2–5w	36.773	1.044	[2.879]	7.439	[2.907]**	1.787	[2.966]
	12–13m	34.273	1.030	[2.550]	0.681	[2.525]	-1.416	[2.493]
Non-agricultural high-skill work	2–5w	1.717	0.028	[.962]	0.512	[.894]	0.431	[1.006]
	12–13m	3.903	-1.539	[1.080]	-1.046	[1.069]	-0.853	[1.145]
Non-agricultural low-skill business	2–5w	8.904	1.607	[1.991]	10.956	[2.339]***	6.214	[2.178]***
	12–13m	8.880	3.465	[1.683]**	1.859	[1.695]	2.960	[1.728]*
Non-agricultural low-skill wage labor	2–5w	16.011	1.537	[2.172]	-2.652	[2.090]	-0.627	[2.200]
	12–13m	15.804	-0.524	[2.242]	-1.684	[2.107]	-2.902	[2.067]
Agricultural work	2–5w	0.700	-0.576	[.401]	-0.501	[.459]	-0.627	[.411]
	12–13m	0.569	-0.293	[.280]	0.060	[.390]	-0.325	[.319]
Illicit work	2–5w	9.281	-3.339	[1.608]**	-2.793	[1.599]*	-3.486	[1.685]**
	12–13m	6.007	0.029	[1.522]	1.590	[1.491]	-0.995	[1.345]

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). Heterosketastic robust standard errors are reported in brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

treatment arm. Therefore post-randomization outcomes are not due to baseline imbalance,

F Survey data validation details

F.1 Methodology

Variable selection

We selected six variables for validation, all with recall periods of two weeks. We chose outcomes with varying degrees of salience (or memorability) and potential social stigma and experimenter bias. We wanted very specific behaviors (e.g. stealing rather than any crime, or marijuana rather than substance abuse). Finally, we wanted sensitive outcomes that were a primary focus of the treatment (stealing) and others that were less so (gambling or expenditures). The variables we selected in the end were:

1. *Stealing.* The survey asked how many times in the last two weeks the respondent stole someone's belongings or deceived or conned someone of money or goods.[20] Based on our fieldwork, we hypothesized that stealing would be the most salient and least socially desirable of all six measures.

---

[20]The survey also measured more serious forms of theft, such as armed robbery, but our qualitative validation focussed on non-violent theft.

Table E.9: Program impacts on baseline outcomes

Outcome	Control mean	ITT regression (N= 947)					
		Therapy only		Cash only		Both	
		ITT	Std. Err.	ITT	Std. Err.	ITT	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Respondents found after 2–5 weeks:							
Antisocial behaviors	0.025	0.088	[.286]	0.175	[.288]	-0.174	[.285]
Economic performance	0.132	-0.321	[.267]	-0.418	[.273]	0.046	[.292]
Time preferences	0.150	-0.230	[.272]	0.089	[.277]	-0.236	[.275]
Self control	-0.002	-0.061	[.239]	-0.225	[.232]	0.064	[.244]
Mental health	-0.019	0.086	[.178]	-0.070	[.185]	-0.106	[.191]
Substance abuse	-0.178	0.353	[.206]*	0.375	[.207]*	0.285	[.210]
Respondents found after 12–13 months:							
Antisocial behaviors	-0.097	0.110	[.279]	0.267	[.282]	0.030	[.282]
Economic performance	0.041	-0.364	[.257]	-0.445	[.260]*	0.013	[.279]
Time preferences	0.141	-0.227	[.270]	0.107	[.272]	-0.177	[.270]
Self control	0.059	-0.073	[.231]	-0.303	[.225]	0.020	[.240]
Mental health	0.014	0.005	[.174]	-0.046	[.182]	-0.055	[.187]
Substance abuse	-0.130	0.309	[.202]	0.342	[.203]*	0.235	[.206]

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). Heterosketastic robust standard errors are reported in brackets.P-values are indicated as follows: *** p<0.01, ** p<0.05, * p<0.1.

2. *Gambling.* The survey asked how many times in the last two weeks the respondent gambled or bet on sports. Beforehand, we hypothesized gambling had a lower level of salience and sensitivity than stealing, but was still somewhat stigmatized.

3. *Marijuana use.* The survey asked how many times in the last two weeks the respondent smoked marijuana. Marijuana use is not socially acceptable across Liberian society overall, but is fairly prevalent in our target demographic. We initially hypothesized underreporting could arise not so much from social stigma but from the discouragement of drug use in the therapy treatment.

4. *Homelessness.* The survey asked how many times in the last two weeks the respondent had to sleep outside, on the street, or in a market stall because they had no other place to sleep or stay. This is a salient variable where we hypothesized respondents might have under-reported from embarrassment or over-reported in order to appear more needy (and eligible for more programs).

5. *Phone charging.* In the expenditure section of the survey, the survey asked how many times in the last two weeks the respondent charged his phone for money. This corresponds to taking one's phone to a kiosk with electricity where one pays a small fee to recharge the battery, a common and routine expense for many Liberians, without stigma and possibly not very memorable. 38% of our sample had a mobile phone at the endline, and 38% reported charging a phone in the last two weeks.

6. *Video Club Attendance.* In the expenditure section of the survey, the survey asked how many times in the last two weeks the respondent went to a video club. These clubs are private businesses where one can go to watch a movie, television show, or football match for a small fee. This is a popular and socially acceptable pastime, as most Liberians do not have electricity or home entertainment. Salience was unclear but likely greater than phone charging.

In part, our use of many non-primary outcomes was deliberate. But, to be frank, our choices were driven more by the practicalities of validation, and in retrospect it would have been useful to focus on more primary outcomes.

Validator staff

Eight local staff performed validations over the two years of data collection. We selected validators from the study's qualitative research staff. These people typically began as survey enumerators, but displayed such skill and rapport with the subjects that we hired and trained them to conduct a separate qualitative research component: longitudinal, formal, open-ended interviews with a different subsample of subjects. All conducted the qualitative validation when they were not working on the formal open-ended interviews.[21]

Each validator received at least 10 days of training on the methods, including both classroom learning and extensive field training. We trained more qualitative researchers than were needed for the exercise. Those who exhibited superior performance during the trainings were selected as

---

[21] All but one were men, and all had a high school education. Two of the men completed roughly half the validations with the remainder doing roughly 10 to 20% each. To find these validators, we trained roughly two to three times the number of people needed from the pool of research staff, selecting only those with the most natural questioning and rapport-building skills for the validation exercise.

validators. The aim of the training was to develop and refine trainees' skills in acquiring informed consent, building rapport with respondents, collecting and recording data, and analytical reasoning. Trainings were held for eight hours each day and, over the course of 10 days, transitioned gradually from exclusive classroom learning to field trainings with short debriefing sessions. Field trainings provided trainees with opportunities to practice the skills and techniques they had learned.

Like any qualitative study, we believe staff recruitment and training to have been among the most important tasks and also the largest start-up cost of this method.

Approach

For each respondent, validators tried to determine whether the respondent had engaged in any of the measured behaviors, even once, in the two weeks preceding the respondent's survey date, as the survey asked about behaviors occurring during the two weeks prior to the survey. We found it optimal for validators to visit each respondent four times, on four separate days, with each visit or "hangout session" lasting approximately three hours. The validator aimed to begin hanging out the day after subjects completed their quantitative surveys and to conduct all four visits in the days following the respondent's endline survey date.

Validators deliberately avoided the feeling of a formal interview and would typically accompany respondents as they went about their business.[22] Validators sometimes took notes during visits, but only in isolated areas out of sight from the respondent.[23] The idea follows from basic principles of ethnography, which seeks to study subjects in their natural settings, similar to those the researcher hopes to generalize about. The intent is to reduce the sense of being in an experimental situation, which ethnographers perceive as creating bias.

The main approach was to engage in casual conversation on a wide range of topics, including the six target topics/measures. The target topics were raised mainly through indirect questions while informally chatting. For example, validators typically started conversations with discussions of family. This was both customary among peers in Liberia and a sign of respect and interest in respondents' lives. It was also a stepping stone for discussing the target behaviors—either because the validator can discuss an issue in their family (someone engaging in one of the activities) or how the respondent's family feels about their current lifestyle and circumstances.

In general, validators found it helpful to tell respondents stories or scenarios about another person or themselves, related to the target measures, then steer the conversation to get information about how respondents had behaved in similar situations, eventually discussing the past two weeks. Validators were careful to present these behaviors and incidents in a non-stigmatized light, for instance by discussing a friend who stole in order to get enough to eat, or how they themselves had periods of homelessness or used drugs and alcohol. Validators found these personal stories (all of which were truthful) and genuineness were essential to building rapport and trust.

Validators might hold these conversations once or twice over the three hours, spending perhaps twenty or thirty minutes in conversation each time, to avoid unnaturally long or awkward conversations. The validator spent the remainder of the three hours in the general vicinity, observing

---

[22]On the first visit validators would obtain verbal consent. We designed the consent script to be informal, and explained that the goal of hanging out with the respondent was to talk about some of the same things they discussed in the survey. In addition to this verbal consent, the formal consent form that preceded the recent survey said that qualitative staff may come and visit them again to gather more information.

[23]e.g. in a toilet stall or teashop. If validators were unable to find a secluded area in which to take notes, they sometimes recorded information in their cell phones, pretending to send a text message.

respondents engaging in their daily activities. This could involve taking a rest in the shade or in a tea shop (as is common) or engaging others in conversation. Validators would also try to talk casually with the respondent's friends, relatives, or neighbors to learn about him (although we considered information from these second-hand sources as insufficient to support a conclusion about the respondents' behaviors, but merely as supporting information).

We found that building a rapport with participants in a short space of time was crucial. To develop trusting and open relationships, validators used techniques, including becoming close to respected local community and street leaders, eating meals together, sharing personal information about themselves, assisting subjects with daily activities, and mirroring participants' appearances and vernacular, as appropriate. In addition, validators tried to maintain neutrality and openness while discussing potentially sensitive topics. For instance, conveying—through stories or otherwise—that illicit behaviors were not perceived negatively, allowed respondents to feel comfortable sharing their involvement in such activities. Validators did not lie to or deceive respondents, however.

Overall, this approach—trust-building, spending time together over the course of several days, assuming the role of an "insider," attempting to obtain admission or discussion of the behavior, clandestine but fairly immediate note-taking, and (as discussed below) close examination of the evidence for each respondent with the investigators—was designed to counter the observer bias and selective recall that concern participant observation.[24] Developing a rapport with respondents, spending time to develop a relationship, and obtaining insider status are considered central to obtaining more honest and valid responses (Baruch, 1981; Bryman, 2003; Fox, 2004). We are not aware of any study, however, that has quantitatively tested this proposition.

Validation sampling and non-response

In each endline survey round we randomly selected study respondents to be validated, stratified by treatment group.[25] Table F.1 describes the samples selected for validation in each survey round over the course of the study. In total, we randomly selected 7.4% of all surveys, 297 in total, for validation.

We found 240 (81%) of the 297.[26] This attrition is an identification concern, but there is little evidence of biased attrition. Excess validation attrition (those who were surveyed but not validated) was not robustly associated with baseline characteristics (see Appendix A.3).

Statistical power

In order to minimize the confidence intervals surrounding any treatment-measurement error correlation, we chose the sample size that maximized the number of interviews we felt qualified validators

---

[24] For general discussions of validity in qualitative methods, see Wilson (1977); LeCompte and Goetz (1982); Power (1989).

[25] For each pair of survey rounds, study participants were randomly divided into blocks (e.g. 1, 2, 3, 4), and block 1 study participants were surveyed before block 2, and block 2 before block 3, etc. Within each block we randomly selected validation subjects using a computer-generated uniform random variable. The selection was performed without replacement in a given pair of survey rounds (e.g. the short-term endline surveys in a given phase), but sampling was performed with replacement across survey rounds. Twenty subjects were validated in more than one round.

[26] We could not find 15 for even the endline survey. We could not validate a further 42 because they were difficult to find even immediately after the survey or (more commonly) because they lived a long distance away. In general, we surveyed respondents who had moved far out of Monrovia, but we were unlikely to validate them because of the time and expense and opportunity cost.

Table F.1: Validation sample, totals and attrition

		Surveys		Validation		Reason for no validation data				
Phase	Round	Target #	Selected	Validated		Unfound at endline	Unfound for validation	% validated (all)	% validated (treatment)	% validated (control)
1	3-week	100	0							
	5-month	100	24	18		2	4	75%	75%	75%
	7-month	100	24	12		1	11	50%	50%	50%
	12-month	100	10	6		3	1	60%	63%	50%
	13-month	100	10	8		2	0	80%	86%	67%
2	3-week	398	26	24		0	2	92%	94%	89%
	5-week	398	27	17		0	10	63%	68%	40%
	12-month	398	28	25		2	1	89%	86%	100%
	13-month	398	44	38		1	5	86%	85%	91%
3	3-week	501	0							
	5-week	501	0							
	12-month	501	35	31		2	2	89%	89%	88%
	13-month	501	69	61		5	3	88%	88%	88%
All		4096	297	240		18	39	81%	81%	80%

Notes: The proportion selected in each round was principally a function of logistical feasibility (e.g. number of available staff), and in some none were selected. As procedures became more familiar and staff more experienced, more could be done over time. The percentage validated in the treatment group includes any treatment (cash, CBT, or both).

could manage logistically.[27] Post hoc calculations of statistical power confirm the estimates we made at the design stage. With a sample of 240, we can detect general over- or under-reporting greater than 17% of the survey mean (14% of the "true" validated mean).[28] Because each treatment arm is a subsample, however, we cannot precisely measure the effect of treatment on misreporting—it is difficult to detect effects greater than 33% of the survey mean (28% of the validated mean). Thus we are principally interested in the sign and magnitude of the treatment effect on misreporting by treatment group.

Coding validated data

Validators were unaware of the respondents' survey responses, and formed their own opinions (based on the evidence collected) about whether respondents engaged in the six activities during the time period captured by the quantitative survey. Every coding recommendation was then discussed with and vetted by one of the authors.

A core part of the validator training included logical reasoning, supporting reasoning with evidence, and writing this down in a clear and structured manner. After each visit, validators made written notes about the relevant data collected, including evidence to support their conclusions, on a standardized form. At the conclusion of the four visits, the validator coded six indicators, one for each behavior, where "1" meant that he had relatively direct evidence that the respondent engaged in the behavior during the recall period, and "0" otherwise.[29]

Validators recorded an average of 1.35 "major" pieces of evidence per respondent per behavior to support their coding decision sheets. This was typically the most persuasive piece or pieces of

---

[27]In general, the validation sample was a balanced subsample of the full sample. Power calculations, based on roughly the first 60 validator interviews, indicated that there was a modest degree of underreporting of all behaviors, sensitive and non-sensitive, but that the correlation between treatment status and measurement error was uncertain—across outcomes it varied in sign and magnitude, but was about zero on average. Thus the chief advantage of maximizing the sample conditional on time available was to shrink the confidence interval to build confidence in our method and the main outcomes of interest. Further validation was mainly limited by the number of validators we felt could be trained and supervised.

[28]We calculated this minimum detectible effect (MDE) using a two-sided hypothesis test with 80% power at a 0.05 significance level, using baseline and block controls when calculating the R-squared 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 .82 in the survey and an MDE of .13 for general over- and under-reporting and .29 for a treatment effect on misreporting. The sensitive behaviors index had a mean of 1.12 in the survey and an MDE of .2 for general over- and under-reporting and .36 for any treatment effect on misreporting. We estimate that doubling the sample size would have increased power by about a third.

[29]Over the course of the exercise, different measures offered different experiences and lessons. Because of its relative frequency and visibility, we suspect marijuana use was the easiest to directly observe. But validators found other behaviors straightforward to discuss in conversation. In the survey and (especially) the validation, phone battery charging led to the most confusion—in particular, did simply charging one's phone count, or did only paying to charge one's phone count? Paid charging was the focus of the survey question (it appeared in an expenditure survey module), but we were concerned that the validators would use a more expansive definition. We attempted to mitigate such differences through trainings and regular discussions on the coding.

Homelessness also proved somewhat challenging to measure and validate, as we discovered its definition is subjective. Circumstances arose that were somewhat ambiguous, such as having no home of one's own but regularly sleeping on a friend's floor or in an acquaintance's market stall. To account for the potential variability in perceptions of homelessness, validators were instructed to include as much information as possible about respondents' living situations in their summary reports. The authors then worked with validators to code a somewhat broad definition of homelessness that included any ambiguous circumstances. Prior to analysis, it was not clear whether survey respondents applied the same definition, and hence we err on the side of finding underreporting in the survey.

evidence rather than all evidence collected.[30] Table F.2 reports evidentiary methods by behavior. In general, the validators used some form of direct or indirect questioning—a direct admission of the behavior or persuasive statements that they did not engage in the behavior. The validators only witnessed or found direct evidence of the behavior in a fifth of cases, or had third party verification in about 6% of cases. In any event, witnessing or third party verification were not sufficient evidence for a final coding. For instance, witnessing had to be followed by questions confirming that the respondent also engaged in the behavior in the two weeks prior to the survey. This accounts for most of the cases where there was more than one piece of evidence highlighted.

In general, the patterns of evidence are fairly commonsensical. Witnessing is limited to observable behaviors such as marijuana, gambling, homelessness, and phone charging. Stories and scenarios where the respondent is invited to comment or discuss are especially common for the most sensitive subject, stealing. Indirect questioning is most common for everyday topics such as homelessness ("Is this your house?") and phone charging ("I need to charge my phone. Where do you usually charge yours?").

Limitations of the approach

While we think, based on our experiences, that this validation exercise gave enough time to gather detailed, accurate information and fostered trust and frankness, there are nonetheless limitations to this approach.

1. POTENTIAL DISRUPTION. The presence, and interactions and conversations with the validators may be intrusive and might disrupt respondents' daily activities, thereby altering the findings. To mitigate this risk, validators wore clothes that would blend in with their respondent's environment, and typically accompanied and assisted respondents in their activities as appropriate (e.g. helping a scrap metal collector scavenge).

2. DIFFERENCES IN RECALL PERIODS. The validation occurred after the time period about which the survey questions had asked, and validators or respondents could have made errors about the relevant window of time (e.g. homelessness could have been observed the week after the survey, and inferred to the time of the survey incorrectly). This is most likely a source of random measurement error.

3. INCONSISTENT QUESTIONS. The survey and validation questions might have been interpreted differently, making it difficult to compare results. As discussed above, phone charging and homelessness proved somewhat difficult to measure consistently. We used close consultations and reviews of the data, and focus groups with survey and validation staff, to maximize consistency.

4. REVERSE HAWTHORNE EFFECT. Training validators to look for certain behaviors could lead them to overreport those behaviors (akin to the problem of "when you have a hammer everything looks like a nail"). This reverse Hawthorne effect would probably be more of a risk if the validation method relied on passive observation. Rather, validation involved active discussion and (usually) a direct admission of the behavior. Also, one of the authors reviewed and discussed the evidence for every subject with the validator.

---

[30]We do not have complete paper records of all evidence collected, and so the 1.35 pieces of evidence is probably an understatement of the full amount of evidence.

Table F.2: Evidentiary methods reported by validators, by behavior

Main evidence techniques	Potentially sensitive behaviors				Expenditures	
	Steal (1)	Marijuana (2)	Gamble (3)	Homeless (4)	Video (5)	Phone (6)
Avg. pieces of evidence	1.1	1.3	1.1	1.7	1.0	1.2
Obs. (All)	240	240	239	240	239	240
Direct question	36%	35%	38%	5%	32%	1%
Indirect question	28%	46%	42%	62%	59%	92%
Story / scenario	36%	6%	13%	12%	2%	1%
Witnessed / found evidence	3%	31%	9%	62%	5%	18%
Third party account	3%	6%	4%	21%	0%	0%
Other / unclear	3%	9%	6%	13%	6%	5%
Obs. (Coded "did not engage" in behavior)	191	118	170	190	93	125
Direct question	38%	44%	39%	5%	34%	0%
Indirect question	26%	46%	44%	60%	58%	98%
Story / scenario	37%	7%	15%	12%	3%	2%
Witnessed / found evidence	2%	3%	1%	65%	2%	1%
Third party account	3%	10%	4%	24%	0%	1%
Other / unclear	2%	1%	1%	14%	4%	0%
Obs. (Coded "did engage" in behavior)	49	122	69	50	146	115
Direct question	29%	25%	36%	4%	30%	2%
Indirect question	33%	46%	38%	70%	60%	86%
Story / scenario	33%	5%	9%	10%	1%	0%
Witnessed / found evidence	10%	59%	28%	52%	7%	37%
Third party account	4%	2%	4%	8%	0%	0%
Other / unclear	8%	17%	17%	6%	8%	10%

Notes: Direct questions imply the validator asked the respondent directly about his engagement in the activity. Indirect questions imply the validator brought up the subject in general conversation (Where do you live? What do you do to make money?). Stories and scenarios are a form of indirect questioning where the respondent is invited to comment. Witnessing or found evidence implies the validator saw the respondent engaging in the activity in question or found physical evidence that the respondent recently engaged in the activity. Third party accounts imply the validator asked the family and friends of the respondent whether or not he engaged in the activity. Other or unclear methods include a handful of cases of unprompted information from the respondent, and also cases where the behavior could be inferred from other knowledge. Mainly it implies that coding was inconclusive or incomplete but is likely a form of questioning.

5. INCREASING SOCIAL DESIRABILITY BIAS. In principle the participant observation method, by building rapport, could lead to a different source of measurement error by (for example) increasing social desirability bias. Our strong sense is that the opposite is true, that trust and rapport reduced the bias, but this is a subjective interpretation and not independently verifiable.

6. CONSISTENCY BIAS. In principle, respondents could recall their survey response and try to remain consistent despite trust-building. This could motivate randomizing the order of validation and survey in the future.

7. NON-BLINDED VALIDATORS. The researcher is not immune from bias in qualitative research (LeCompte and Goetz, 1982; LeCompte, 1987). We are especially concerned with any bias correlated with treatment. While validators weren't given the subject's treatment status, it's possible and even likely that this could come up during the extended conversations. Thus there is a danger that the validators' biases will be correlated with treatment. The trust-building and preference for direct admission of the behavior was intended to mitigate this risk, but it still remains.

Most importantly, it seems unlikely that validators would commit most of these errors differentially across study arms. Misreporting correlated with treatment is still a risk under the consistency bias and non-blinded limitations, but the in-depth focus on a handful of questions, time invested, and trust-building is designed to counteract these biases as much as possible. If so, the qualitative validation method may be most useful at building confidence estimated treatment effects.

Finally, like any qualitative work, this is not an off-the-shelf tool. To select and refine the variables, recruit and train validators, and monitor quality of the data requires the researcher to have some familiarity with the context and population and at least basic experience in qualitative data collection.

Replicability of the approach

There are three reasons to think that this method could be replicated in other developing country field experiments and observational analysis using surveys. First, the expertise needed to implement the method effectively exists in most countries. Indeed, it should be considerably simpler to implement outside than inside Liberia. After fourteen years of civil war, and with one of the lowest human development indices in the world, Liberia has very low local research capacity, even compared to other poor and post-conflict states.

Second, most social scientists are nearly as well prepared to design and implement the approach as they are a new survey instrument or measure. Like any measure or method, it takes local knowledge, care, and extensive pretesting to develop a credible approach, and can benefit from someone with expertise in the subject area. In our case, one of the field research managers had some background in qualitative work and quality assurance, which we believe improved the quality of training and selection of the validator staff.

Third, the cost of the data collection is not necessarily large relative to many field experiments or large-scale panel surveys. In this instance, the fixed cost of startup was primarily in the recruitment and training of the small number of validators—approximately 2 to 3 weeks of work. We estimate

the marginal cost of validation was roughly \$80 per respondent, mainly in wages and transport. By comparison, the marginal cost of surveying a respondent was roughly \$70.[31]

While this method is considerably more expensive than survey experiments, it is more in line with the depth and cost of commonplace efforts to improve consumption measurement through the use of diaries physical measurement.[32] For crucial measures in large program evaluations, or for statistics informing major policies, the cost is small relative to the intervention, larger study, or larger purpose. For instance, as a proportion of total expenditures on the study, this validation exercise cost under 3% of all research-related costs, and less than 1–2% of program plus research costs.

F.2 Further analysis

Misreporting levels

Table F.3 reports our proxy of survey over-reporting: the simple survey-validation differences, with p-values from a t-test of the difference from zero. Negative values indicate survey under-reporting, assuming the validator measure is more accurate of course. As noted above, we have the statistical power to detect differences greater than about 17% of the survey mean.

Overall, gambling seems to be slightly underreported in every treatment arm, and highly under-reported by men in the control and cash only groups. For instance, 33% of the cash only group admitted to gambling during validation, compared to 13% during the survey. Some of this underreporting could be due to ambiguous behaviors being coded as gambling in validation interviews but not in the survey. But the fact that underreporting is smaller in the therapy arms suggests that the underreporting is not an artifact of different definitions, but rather reflects a strategic response to treatment status.

If we look at stealing, marijuana use, and homelessness, however, none of the survey-validation differences are statistically significant. There is possibly some slight underreporting of drug use and slight over-reporting of stealing, but the magnitudes are generally small in the sense that they are less than 10% of the survey means reported in Table 9. The sample size is small, however, and so many of these differences are not precisely estimated.

We see much stronger evidence of underreporting of expenditures in the survey. The difference for both expenditures is -0.27 in the full sample (Table F.3, Column 6). This difference is large—about a third of the survey mean reported in Table 9. Expenditure underreporting is largest for the video club measure, but both expenditures appear to be underreported. Interestingly, the mean differences appear to be smaller and less statistically significant if the men received one of the treatments. We return to these differences across treatment arms below.

Patterns of survey under- and over-reporting

In our validation exercise, there may be cases where the validation technique did not report a behavior that was reported in the survey. Table F.4 reports the number of cases where the survey

---

[31]Both figures were driven by the fact that it typically took one to two days of searching to find each respondent for surveying, plus the time to survey itself. Both surveying and validating in Liberia were expensive by the standards of household surveys, largely because of the cost of operating in a fragile, post-conflict state and the great difficulties in tracking such an unstable population.

[32]In one extreme example, in the India NSS consumption survey, enumerators physically measure the volume of all food consumption Group (2003).

Table F.3: Survey over-reporting, estimated by the mean difference between survey and validation measures (y^Δ)

	Potentially sensitive behaviors					Expenditures		
	All (0-4)	Steal	Marijuana	Gamble	Homeless	All (0-2)	Video	Phone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full sample	-0.10	0.02	-0.03	-0.11	0.02	-0.27	-0.19	-0.08
	0.17	*0.57*	*0.24*	***0.00***	*0.45*	***0.00***	***0.00***	***0.00***
Control group	-0.07	0.03	-0.02	-0.12	0.03	-0.50	-0.29	-0.22
	0.64	*0.57*	*0.71*	*0.09*	*0.60*	***0.00***	***0.00***	***0.00***
Therapy only	-0.04	0.02	0.00	-0.07	0.02	-0.17	-0.13	-0.04
	0.80	*0.77*	*1.00*	*0.29*	*0.77*	*0.08*	*0.07*	*0.53*
Cash only	-0.29	-0.02	-0.05	-0.20	-0.03	-0.23	-0.18	-0.05
	0.04	*0.80*	*0.37*	***0.00***	*0.42*	***0.03***	***0.03***	*0.32*
Therapy + cash	0.02	0.03	-0.05	-0.03	0.06	-0.19	-0.16	-0.03
	0.91	*0.57*	*0.37*	*0.66*	*0.25*	***0.01***	***0.02***	*0.48*
Observations	239	238	238	238	239	239	238	239

Notes: Columns 1 to 8 report the simple mean differences in the survey and validation measures for the full sample and for each treatment arm, along with p-values for as t-test of whether the mean is different from zero. We bold p-values ≤ 0.05.

and validation measures do not agree, divided into cases of survey over- and under-reporting relative to the validation measure. Over-reporting is driven by stealing, gambling, homelessness and going to the video store. Over-reporting is limited for marijuana use and phone charging, which are some of the least ambiguous and most habitual activities.

Another way to understand this point is to rerun equation 3 in the paper but omit block fixed effects and restrict $\tilde{\beta}_1 = 0$ and $\tilde{\beta}_3 = 0$:

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

In this case, $\tilde{\beta}_0$ is an estimate of survey over-reporting. Table F.5 reports these results, Panel (a) with the restrictions $\tilde{\beta}_1 = 0$ and $\tilde{\beta}_3 = 0$ and Panel (b) without, for comparison. Looking at the sensitive behaviors in Panel (a), we see evidence of survey overreporting ranging from rough 12

Table F.4: Survey under-reporting and over-reporting by behavior

	Sensitive behaviors					Expenditures			
	All	Steal	Marijuana	Gamble	Homeless	All	Video	Phone	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				A. UNDERREPORTS					
Total count	109	23	21	46	19	99	68	31	208
				B. OVER-REPORTS					
Total count	86	27	14	21	24	34	23	11	120

to 15%. Moreover, $\tilde{\beta}_0$ and $\tilde{\beta}_2$ are relatively similar in both Panels (a) and (b), suggesting that treatment has little effect on this survey over-reporting.

We do not know for certain why up to 15% of people would report a sensitive behavior in the survey but not in the validation exercise but there are several plausible explanations. First, survey respondents may not have considered the "last two weeks" recall period carefully, and reported behavior over a wider range. Validators were trained to be more strict with the recall window. Second, although we tried our best to maintain consistent definitions across the survey and the validation exercise, validators might have used more restrictive definitions of the behavior in question. Finally, validators may simply have been more conservative in their coding of these behaviors, or set too high a bar for certainty.

They key, however, is that there is no evidence that misreporting is associated with treatment status–which itself is the core finding from the general analysis of the validation exercise.

Treatment effects in the overall sample versus the validation sample

In this section we investigate whether the treatment effects observed in the validation sample are similar to those observed in the validation sample. Panel (a) of Table F.6 takes the survey measures of our six validated outcomes, and reports ITT estimates in the validated sample (N=238) and the full sample. Panel (b) takes the validator measures of our six outcomes, and reports ITT estimates in the validated sample (N=238). Although the validation sample only has 238 observations, and so standard errors are large, the estimate treatment effects are qualitatively similar across all three sets of regressions.

F.3   Adjusted treatment effects

We estimate the effect of each treatment on survey over-reporting, in Table F.7. These estimates effectively take the simple survey-validation differences in Panel A of Table 10 and estimate the difference across treatment arms, adjusting for baseline covariates as well as block fixed effects. We use these to calculate an adjusted treatment effect.

First, the results imply that the adjusted treatment effect of therapy and cash on sensitive behaviors overall is no lower than what we estimate with self-reported survey data, and may even be larger (Column 1). This holds true for each of the individual sensitive behaviors, save marijuana use. Despite the large standard errors introduced by the small validation sample, the adjusted treatment effect on all sensitive behaviors is larger and significant at the 1% level.

Meanwhile, the underreporting of gambling does not have a statistically significant association with treatment. Those who received cash alone underreported gambling to the surveyors more often than control group members, and so the measurement error in gambling is probably a combination of a general desirability bias as well as one correlated with treatments. A larger sample size would be needed to separate these more precisely.

In contrast, the slight underreporting of expenditures behaviors in the survey (seen in Table F.3 above) implies that the short term increase in survey-based expenditures due to cash could be due to measurement error correlated with treatment. The adjusted treatment effect of therapy plus cash is generally negative but not statistically significant (Column 6). We see a similar pattern with another expenditure-related item, homelessness, in Table F.7—the survey-reported decline in homelessness tends to disappear with adjustment.

Table F.5: Estimates of the correlation between treatment and measurement error, omitting block fixed effects

(a) Constrained regression, $\tilde{\beta}_1 = 0$ and $\tilde{\beta}_3 = 0$

Covariate	Stealing	Marijuana	Gambling	Homeless	All sensitive (0–4)	Video Club	Phone Charging	All expenditures (0-2)
$\hat{\beta}_0$ (Constant)	0.143 [.025]***	0.121 [.029]***	0.124 [.025]***	0.127 [.026]***	0.456 [.078]***	0.250 [.045]***	0.088 [.025]***	0.278 [.066]***
$\hat{\beta}_2(y^v)$	0.388 [.077]***	0.707 [.046]***	0.209 [.061]***	0.493 [.074]***	0.545 [.050]***	0.284 [.060]***	0.640 [.049]***	0.495 [.056]***

(b) Unconstrained regression

Covariate	Stealing	Marijuana	Gambling	Homeless	All sensitive (0–4)	Video Club	Phone Charging	All expenditures (0-2)
$\tilde{\beta}_0$ (Constant)	0.156 [.052]***	0.100 [.053]*	0.128 [.053]**	0.174 [.061]***	0.373 [.171]**	0.350 [.106]***	0.036 [.035]	0.296 [.173]*
$\tilde{\beta}_1$								
Therapy	-0.022 [.074]	0.007 [.079]	-0.006 [.074]	-0.034 [.081]	0.113 [.228]	-0.190 [.128]	0.098 [.072]	-0.046 [.205]
Cash	-0.010 [.071]	0.043 [.083]	-0.079 [.063]	-0.131 [.068]*	-0.078 [.216]	-0.064 [.138]	0.053 [.060]	0.078 [.215]
Both	-0.018 [.069]	0.033 [.080]	0.059 [.077]	-0.023 [.082]	0.286 [.234]	-0.139 [.144]	0.055 [.061]	-0.167 [.203]
$\tilde{\beta}_2(y^v)$	0.487 [.159]***	0.767 [.091]***	0.300 [.107]***	0.398 [.152]***	0.666 [.105]***	0.035 [.126]	0.527 [.093]***	0.327 [.130]**
$\tilde{\beta}_3$								
Therapy×y^v	-0.176 [.236]	0.011 [.125]	-0.115 [.175]	0.008 [.220]	-0.145 [.144]	0.426 [.173]**	0.090 [.143]	0.248 [.169]
Cash×y^v	-0.205 [.212]	-0.122 [.132]	-0.049 [.150]	0.293 [.193]	-0.109 [.143]	0.121 [.174]	0.171 [.132]	0.073 [.167]
Both×y^v	-0.041 [.210]	-0.112 [.128]	-0.221 [.181]	0.052 [.225]	-0.245 [.135]*	0.437 [.174]**	0.216 [.126]*	0.401 [.164]**

(c) Constrained regression, $\tilde{\beta}_2 = 0$ and $\tilde{\beta}_3 = 0$

	Dependent variable (N=239)							
	$y^s - y^v$, Sensitive behaviors					$y^s - y^v$, Expenditures		
Covariate	Stealing	Marijuana	Gambling	Homeless	All (0–4)	Video Club	Phone Charging	All (0–2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_o (Constant)	0.034 [.063]	-0.017 [.037]	-0.117 [.066]*	0.033 [.067]	-0.067 [.139]	-0.288 [.081]***	-0.217 [.061]***	-0.500 [.094]***
β_1								
Therapy	-0.015 [.089]	0.017 [.059]	0.043 [.095]	-0.015 [.090]	0.030 [.203]	0.159 [.105]	0.180 [.085]**	0.333 [.133]**
Cash	-0.050 [.089]	-0.033 [.066]	-0.080 [.089]	-0.066 [.078]	-0.224 [.195]	0.111 [.115]	0.168 [.078]**	0.274 [.138]**
Both	-0.002 [.083]	-0.031 [.064]	0.085 [.094]	0.030 [.089]	0.083 [.187]	0.129 [.103]	0.185 [.076]**	0.310 [.116]***

Table F.6: Correlation between treatment and responses

	Intent to treat estimates											
	Panel a: Dependent variable: Survey response (y^s)											
	Validated sample (N=238)						Full sample (N=3764)					
	Therapy		Cash		Both		Therapy		Cash		Both	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sensitive	-0.226	[.223]	-0.257	[.224]	-0.123	[.212]	-0.184	[.092]**	-0.056	[.095]	-0.398	[.090]***
Stealing	-0.080	[.080]	-0.057	[.077]	-0.047	[.075]	-0.045	[.028]	-0.031	[.029]	-0.106	[.027]***
Marijuana	-0.028	[.092]	-0.022	[.093]	-0.008	[.097]	-0.023	[.040]	0.012	[.040]	-0.068	[.040]*
Gambling	-0.072	[.075]	-0.109	[.071]	-0.027	[.075]	-0.086	[.026]***	0.025	[.029]	-0.099	[.026]***
Homeless	-0.049	[.082]	-0.065	[.079]	-0.045	[.082]	-0.030	[.031]	-0.061	[.031]**	-0.125	[.029]***
Expenditures	0.145	[.141]	0.098	[.133]	0.287	[.140]**	0.000	[.050]	0.078	[.052]	0.075	[.050]
Video	0.046	[.090]	0.005	[.089]	0.165	[.088]*	0.001	[.032]	0.026	[.034]	0.011	[.033]
Phone	0.099	[.095]	0.093	[.094]	0.122	[.098]	-0.002	[.035]	0.052	[.037]	0.064	[.035]*

	Panel b: Dependent variable: Validated response (y^v)					
	Validated sample (N=238)					
Sensitive	-0.199	[.226]	0.006	[.235]	-0.180	[.226]
Stealing	-0.052	[.078]	-0.010	[.079]	-0.032	[.080]
Marijuana	-0.034	[.094]	0.023	[.097]	0.025	[.100]
Gambling	-0.092	[.089]	-0.016	[.091]	-0.100	[.087]
Homeless	-0.021	[.079]	0.016	[.079]	-0.073	[.075]
Expenditures	-0.192	[.140]	-0.169	[.137]	-0.015	[.134]
Video	-0.122	[.089]	-0.102	[.086]	0.040	[.083]
Phone	-0.082	[.097]	-0.080	[.093]	-0.066	[.098]

Notes: Treatment effects are estimated via OLS controlling for block fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table F.7: Estimates of adjusted treatment effects by outcome and treatment

| | Potentially sensitive behaviors | | | | | Expenditures | | |
| | All (0-4) | Steal | Marijuana | Gamble | Homeless | All (0-2) | Video | Phone |
Adjusted ATE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cash only	0.121	-0.014	0.058	0.072	0.001	-0.205	-0.079	-0.132
	[.196]	[.088]	[.074]	[.093]	[.077]	[.143]	[.117]	[.079]*
Therapy only	-0.190	-0.041	-0.031	-0.131	0.007	-0.334	-0.167	-0.172
	[.202]	[.082]	[.066]	[.104]	[.087]	[.148]**	[.121]	[.090]*
Therapy and cash	-0.516	-0.122	-0.048	-0.198	-0.156	-0.239	-0.125	-0.120
	[.194]***	[.075]	[.071]	[.095]**	[.082]*	[.137]*	[.113]	[.077]
Observations, survey/validation	3765 / 239	3764 / 238	3762 / 238	3763 / 238	3765 / 239	3763 / 239	3761 / 238	3759 / 239

Notes: The survey-based ATE estimates pool all survey rounds and regress each outcome on treatment indicators and block fixed effects. Standard errors are robust and clustered by individual. Estimates of the bias from treatment come from a regression of the difference in the survey and validation measures on an indicator for treatment arms, controlling for block fixed effects and each endline round. Standard errors are robust and clustered by block. The difference is an estimate of the true treatment effect after adjusting for observed bias. It is calculated as the linear difference of the estimates and the standard error is calculated via bootstrapping. *** p<0.01, ** p<0.05, * p<0.1

References

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Aron, A. (2007). The neural basis of inhibition in cognitive control. *Neuroscientist 13*(3), 214–28.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of adhd. *Psychological Bulletin 121*(1), 65–94.

Baruch, G. (1981). Moral tales: parents' stories of encounters with the health professions. *Sociology of Health & Illness 3*(3), 275–295.

Blattman, C. and J. Annan (2015). Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *forthcoming in American Political Science Review*.

Blattman, C., N. Fiala, and S. Martinez (2014). Generating skilled employment in developing countries: Experimental evidence from Uganda. *Quarterly Journal of Economics 129*(2), 697–752.

Bryman, A. (2003). *Quantity and quality in social research*. New York: Routledge.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Draca, M. and S. Machin (2015). Crime and Economic Incentives. *Annual Review of Economics 7*, 389–408.

Eigsti, I.-M., V. Zayas, W. Mischel, Y. Shoda, O. Ayduk, M. Dadlani, M. Davidson, A. J. Lawrence, and B. Casey (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science 17*(6), 478–84.

Ersche, K., A. Turston, S. Pradhan, E. Bullmore, and T. Robbins (2010). Drug addiction endophenotypes: impulsive versus sensation-seeking personality traits. *Biological Psychiatry 68*(8), 770–3.

Fafchamps, M., D. J. McKenzie, S. Quinn, and C. Woodruff (2014). When is capital enough to get female microenterprises growing? Evidence from a randomized experiment in Ghana. *Journal of Development Economics 106*(1), 211–226.

Fox, R. C. (2004). Observations and reflections of a perpetual fieldworker. *The ANNALS of the American Academy of Political and Social Science 595*(1), 309–326.

Group, N. E. (2003). Suitability of different reference periods for measuring household consumption. results in pilot survey. *Economic and Political Weekly 38*(4), 25–31.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Jamison, J. and D. Karlan (2011). Measuring preferences and predicting outcomes. *Working paper*.

Judge, T. A. and J. E. Bono (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of applied Psychology 86*(1), 80.

Judge, T. A., A. Erez, J. E. Bono, and C. 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.

Karlan, D., R. Knight, and C. Udry (2015). Consulting and capital experiments with microenterprise tailors in ghana. *Journal of Economic Behavior and Organization 118*, 281–302.

LeCompte, M. D. (1987). Bias in the biography: Bias and subjectivity in ethnographic research. *Anthropology & Education Quarterly 18*(1), 43–52.

LeCompte, M. D. and J. P. Goetz (1982). Problems of reliability and validity in ethnographic research. *Review of educational research 52*(1), 31–60.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *American Economic Review 80*(2), 319–323.

Nelson, C. A. e. a. (2007). Cognitive recovery in socially deprived young children: The bucharest early interventino project. *Science 318*(5858), 1937–1940.

Paluck, E. L. (2009). What's in a norm? Sources and processes of norm change. *Journal of Personality and Social Psychology 96*, 594–600.

Power, R. (1989). Participant observation and its place in the study of illicit drug abuse. *British Journal of Addiction 84*(1), 43–52.

Udry, C. (2010). The Economics of Agriculture in Africa: Notes on a Research Program. *African Journal of Agricultural and Resource Economics 5*(1), 284–299.

Wilson, S. (1977). The use of ethnographic techniques in educational research. *Review of educational research 47*(1), 245–265.