Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes

James J. Heckman
University of Chicago

Lena Malofeeva
University of Arizona

Rodrigo Pinto
University of Chicago

Peter Savelyev
University of Chicago

Measuring Education Outcomes:
Moving from Enrollment to Learning
Center for Universal Education at Brookings
Washington, DC
June 2, 2010

All figures in this presentation are based on the authors' own calculations

Acknowledgments

This research was supported by the Committee for Economic Development through a grant from the Pew Charitable Trusts and the Partnership for America's Economic Success; Susan Thompson Buffett Foundation; Mr. Robert Dugger; and NICHD (R01HD043411). The views expressed in this presentation are those of the authors and not necessarily those of the funders listed here. We are grateful to Christopher Hansman, Kegon Tan Teng Kok, Min Ju Lee, Xiliang Lin and Yun Pei for their excellent research assistance.

Overview

- Noncognitive skills deserve as much attention of policymakers as cognitive skills
- I will show that noncognitive skills:
  1. are malleable
  2. affect life outcomes
- Our evidence is based on a randomized experiment
- Perry Preschool Program:
  1. the most influential early education experiment
  2. the earliest experiment of the kind

Overview: Perry Preschool Program

- **Program Type:** Randomized Experiment (123 children, 58 Treated, 65 Controls)
- **Age:** Children were 3–5 while treated
- **Place and Time:** Conducted in Ypsilanti, Michigan, in the early 1960s
- **Target Population:** Low IQ ($IQ \leq 85$), Low SES, disadvantaged African-American Children
- **Data:** Multiple measurements at ages 3–15, 19, 27 and 40

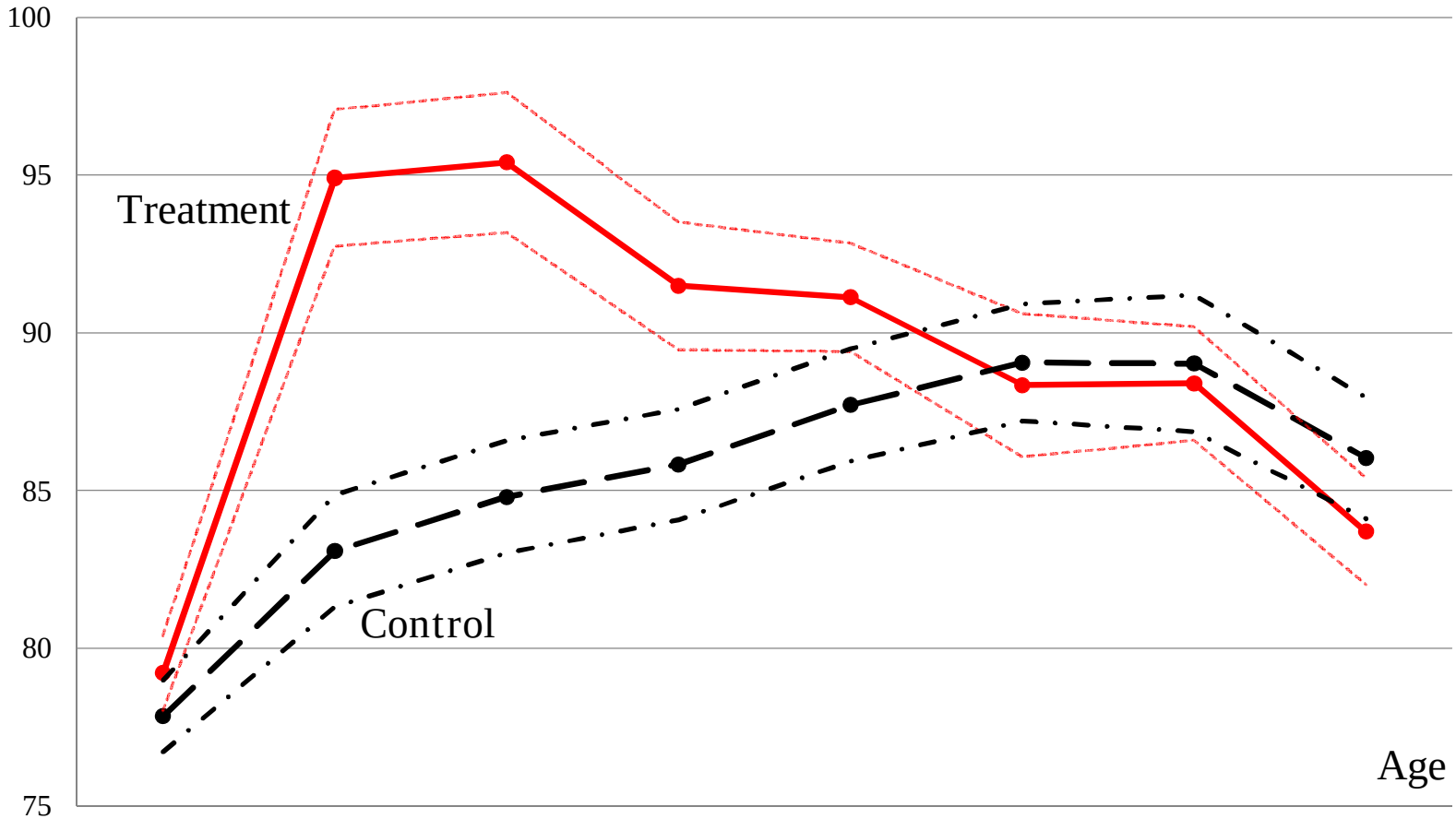
Overview: Perry Preschool Curriculum

- conducted for 2.5 hours per day, five days per week during the school year, supplemented by weekly home visits from teachers
- based on *active participatory learning*, in which children and adults were seen as equal partners in the learning process
- promoted abilities to plan, execute, and evaluate tasks
- emphasized teaching self-control, planning in social settings, and resolution of interpersonal conflicts

Overview: The Perry IQ Pattern

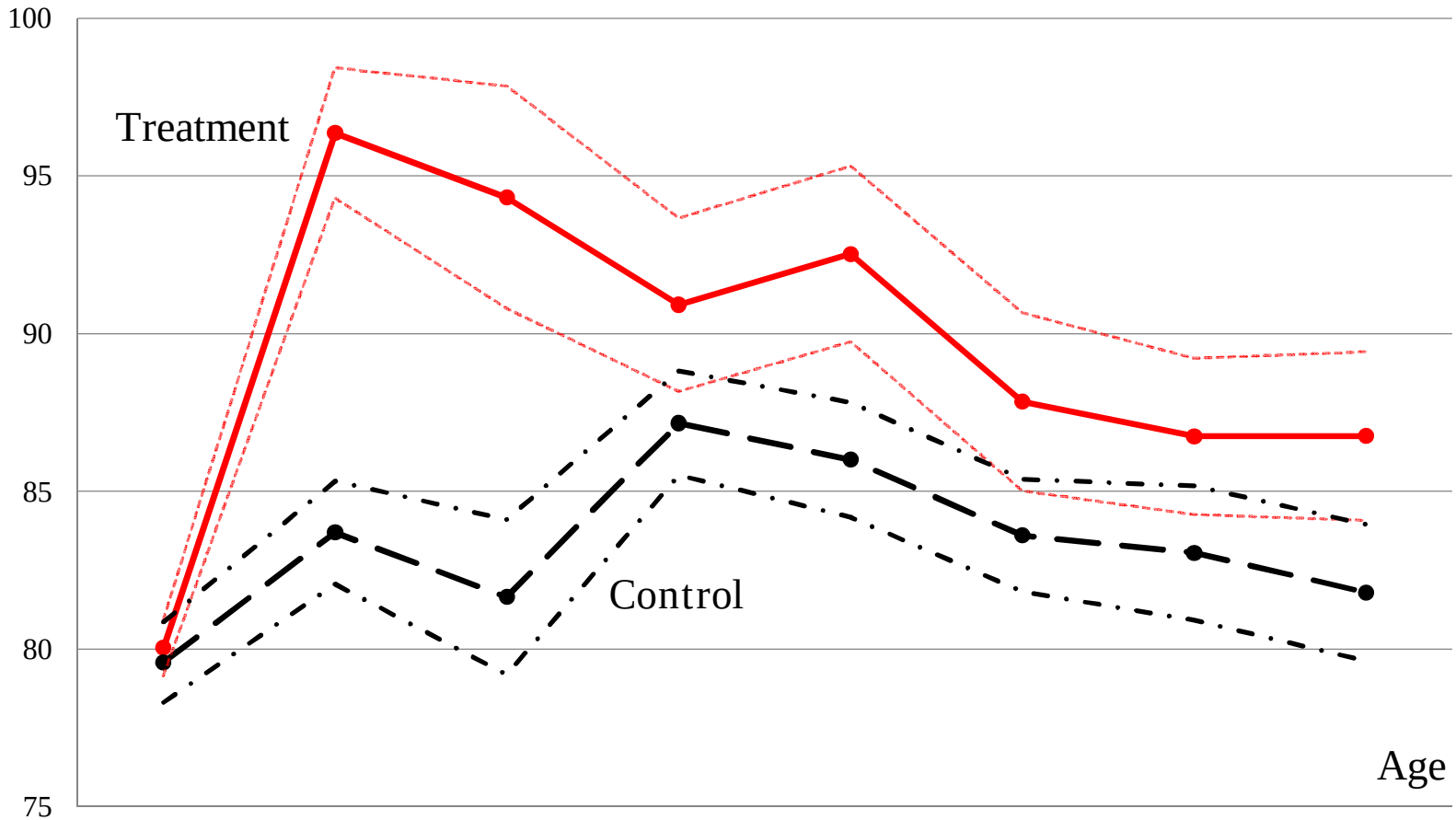
- The Perry program initially boosed IQ, but the effect quickly faded after the end of the two-year program.
- Yet statistically significant treatment effects on lifetime outcomes up to age 40 were observed for both gender groups, even after accounting for the multiplicity of similar variables and for the compromised randomization (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010b).
- The rate of return is high (7-10%) and statistically significantly different from zero (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010a).
- The cost-benefit ratios are large and precisely determined (Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010a).
- IQ was not raised. Traits other than cognition are involved, which we confirm in this paper.

Figure 1: Stanford-Binet IQs, Perry Males



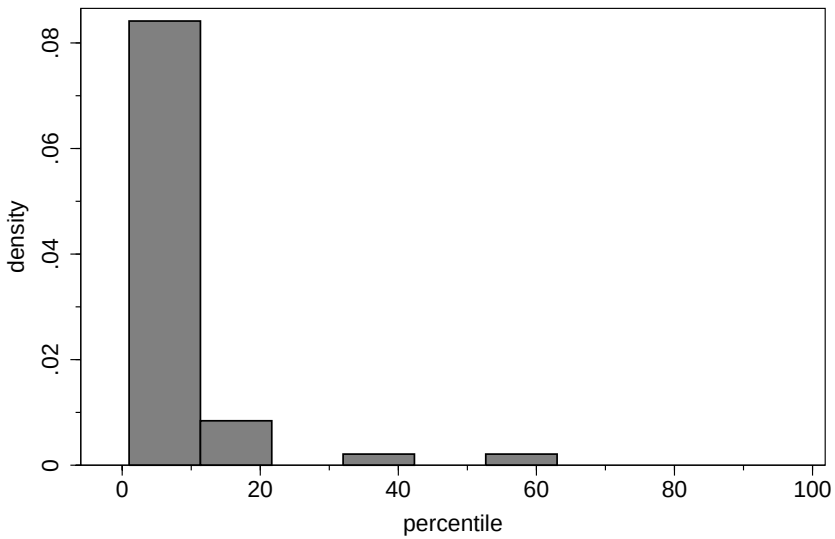
	Entry	4	5	6	7	8	9	10
Treatment Mean	79.21	94.91	95.40	91.48	91.12	88.33	88.39	83.70
Control Mean	77.85	83.08	84.79	85.82	87.71	89.05	89.03	86.03

Figure 2: Stanford-Binet IQs, Perry Females

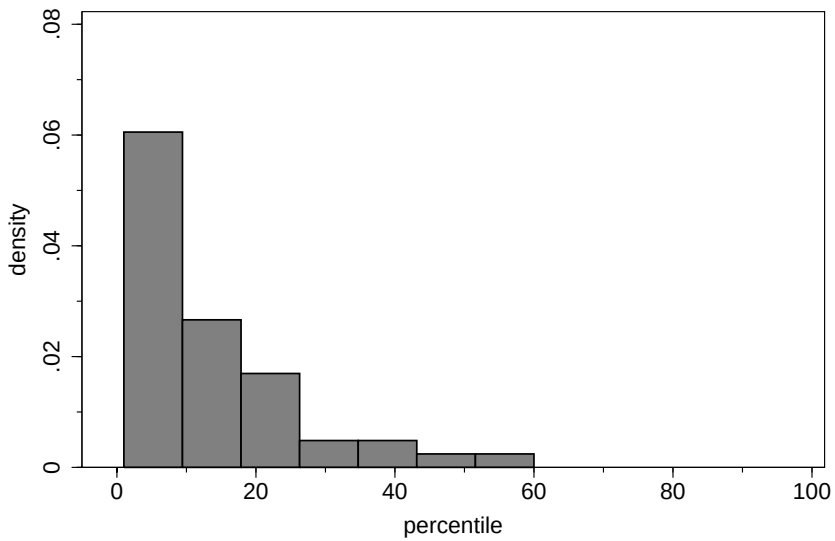


	Entry	4	5	6	7	8	9	10
Treatment Mean	80.04	96.36	94.32	90.91	92.52	87.84	86.74	86.75
Control Mean	79.58	83.69	81.65	87.16	86.00	83.60	83.04	81.79

Figure 3: CAT Score, Both Genders, Age 14



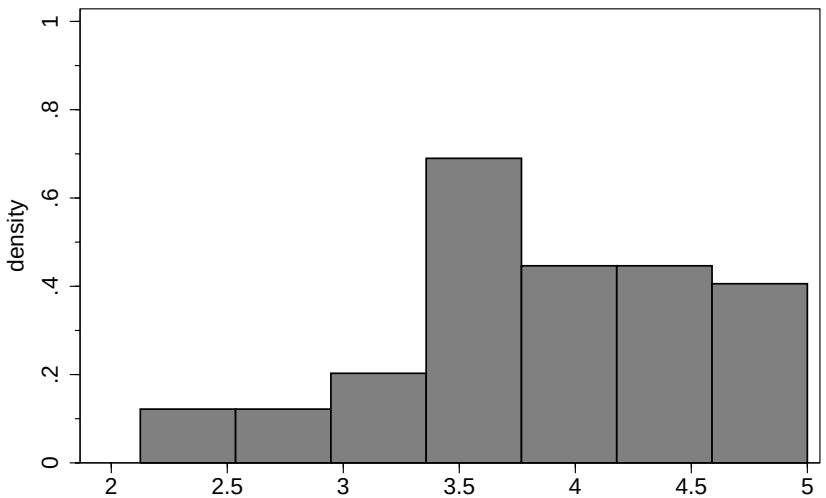
(a) Control



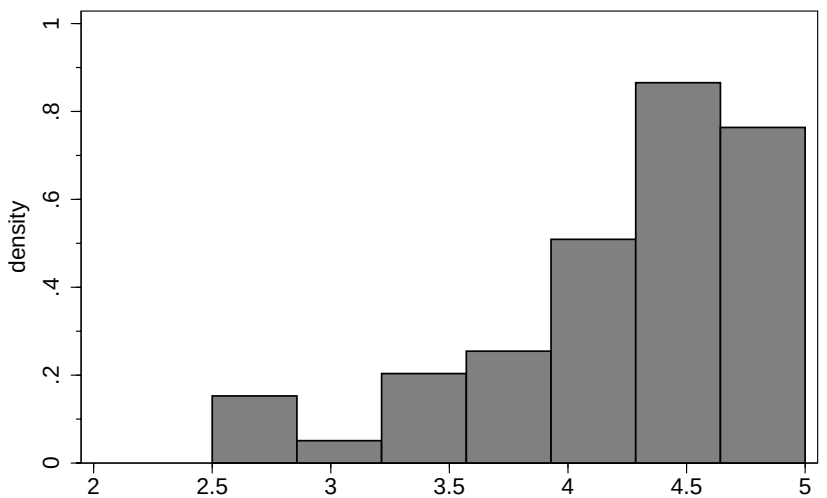
(b) Treatment

CAT is California Achievement Test. Achievement Tests are loaded on noncognitive skills (Borghans, Golsteyn, and Heckman, 2009). One-sided p-value for the difference in means is 0.001.

Figure 4: Personal Behavior Index, Both Genders



(a) Control



(b) Treatment

- Treatment shifts the distribution upwards (1=bad;...;5=good).
- Statistically significant treatment effect is observed: $p = 0.002$.
- The Personal Behavior Index is an unweighted average of four items: "absences and truancies", "lying or cheating", "steals" and "swears or uses obscene words".

Overview: Three Questions

1. How do cognitive and noncognitive traits change in response to the experimental invention?
   - Noncognitive traits positively affected for both genders
   - Cognitive trait change only for females
2. How do the experimentally-induced changes in traits translate into life outcomes?
   - Males: strong mediation of "Personal behavior" only
   - Females: multiple traits are mediators
3. How do treatment effects differ across quantiles of outcome distributions? Do those with initially higher levels of traits benefit the most of the least?
   - Effects on achievement tests and education are stronger for better achievers
   - Effects on arrests are stronger for people with propensity to commit more crimes

Measures of Traits

Our Measures of Cognitive and Noncognitive Traits

- Human capacities are legion and, in general, they cannot be reduced to a scalar human capital.
- A large literature establishes the importance of cognition (Gottfredson, 1997; Jensen, 1998).
- Noncognitive traits are not as well established as cognitive (Borghans et al., 2008).
- We use two noncognitive measures chosen based on AIC information criterion: Personal Behavior and Socio-Emotional State.

Table 1: Description of Measures

Traits	# of Measures	Age	Description of Measures
Cognition	3	7	Stanford-Binet IQ
		8	Stanford-Binet IQ
		9	Stanford-Binet IQ
Personal Behavior	4	7-9	PBI absences or truancies
		7-9	PBI lying or cheating
		7-9	PBI steals
		7-9	PBI swears or uses obscene words
Socio-Emotional State	4	7-9	PBI appears depressed
		7-9	PBI withdrawn and uncommunicative
		7-9	PBI friendly and well-received by other pupils
		7-9	PBI appears generally happy

Empirical Results

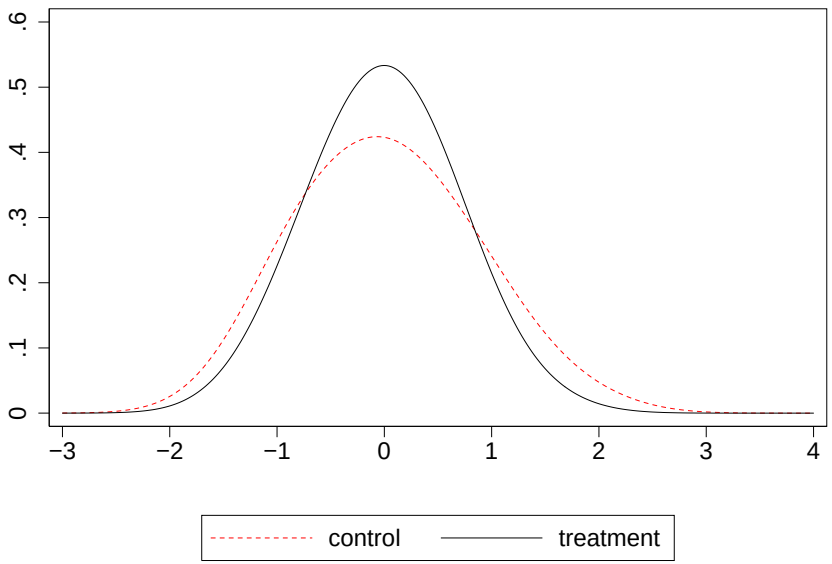
Three empirical analyses.

1. We study how treatment affects the marginal distributions of cognitive and noncognitive traits.
2. We decompose mean treatment effects on outcomes into components corresponding to changes in each trait.
3. We split distributions of outcomes into quantiles and compare treatment effects calculated for different quantiles.

1. The Effect of the Perry Program on Capabilities

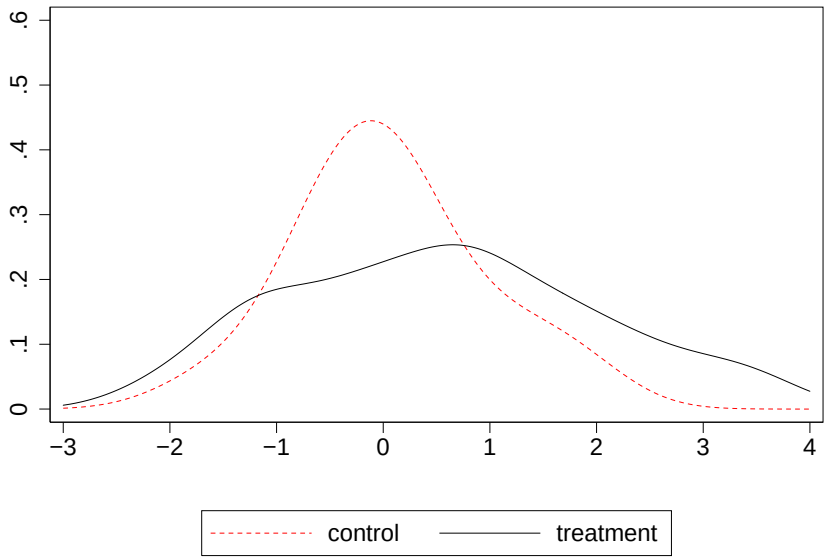
Figure 5: Cognition Factor Scores, Kernel Density

$p = .427$



(a) Males

$p = .095$

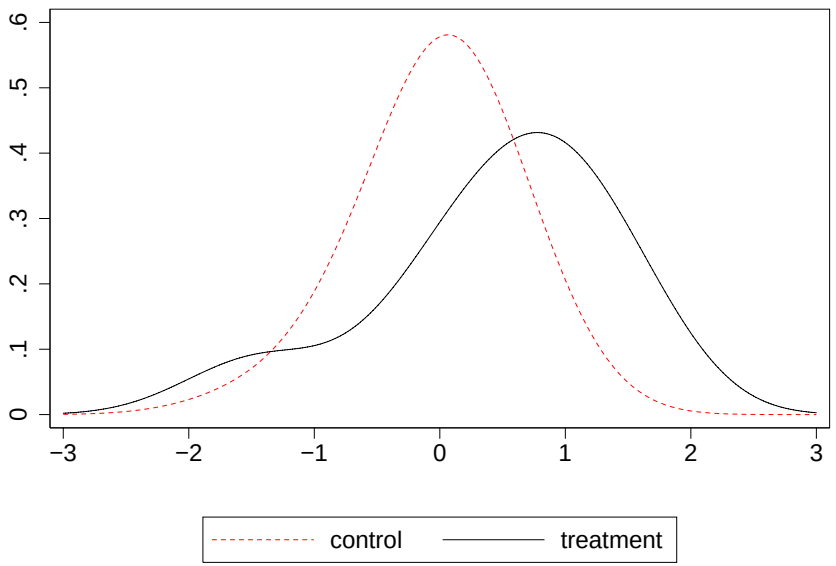


(b) Females

- Cognition enhanced for females, but not for males

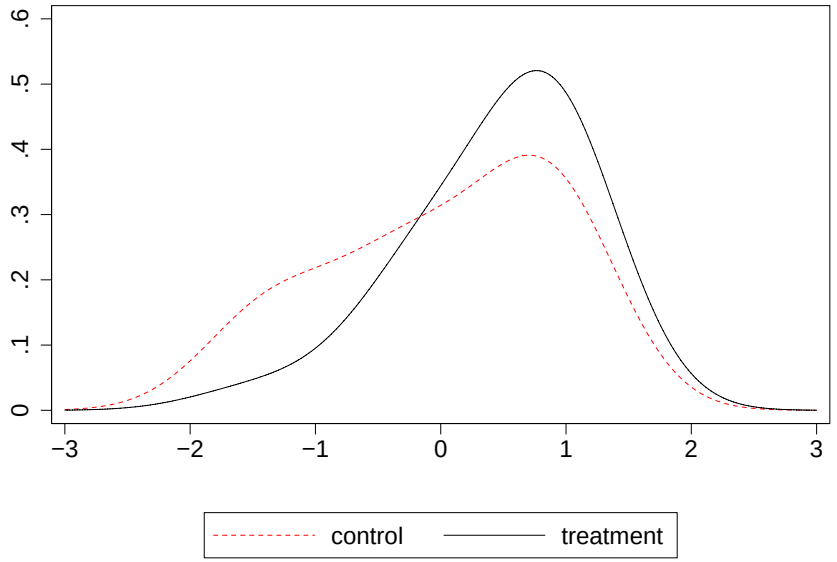
Figure 6: Personal Behavior Factor Scores, Kernel Density

$p = .005$



(a) Males

$p = .042$

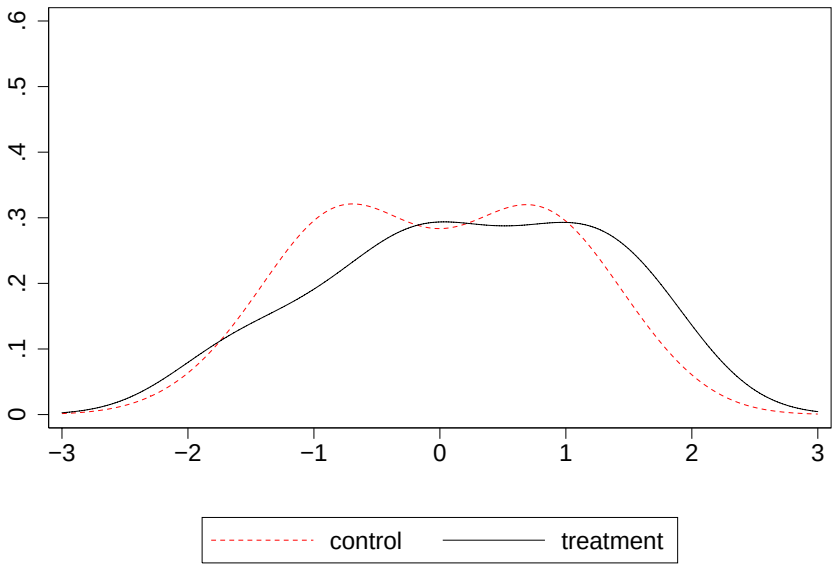


(b) Females

- Personal Behavior enhanced for both genders, with the effect for for males being especially strong
- Elimination of low traits for females (standard result)
- Enhancement of all levels of traits for males

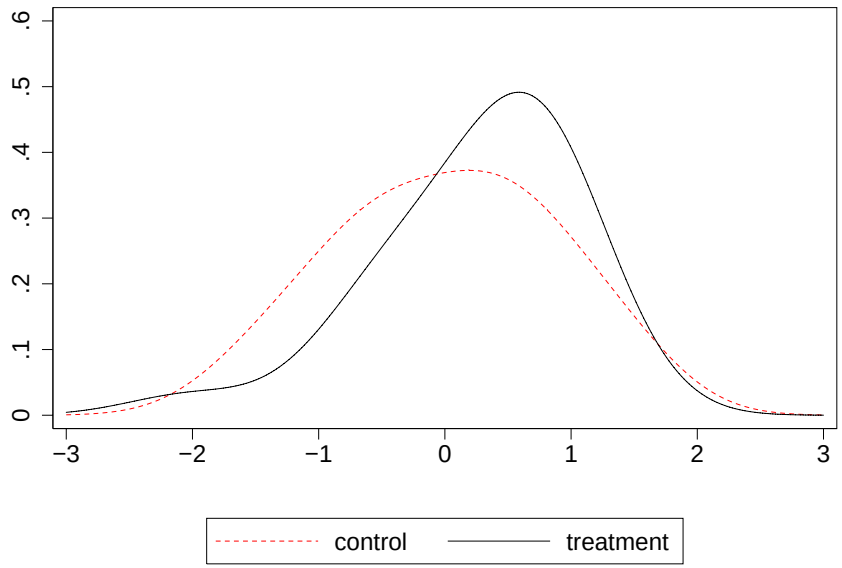
Figure 7: Socio-Emotional State Factor Scores, Kernel Density

$p = .177$



(a) Males

$p = .151$

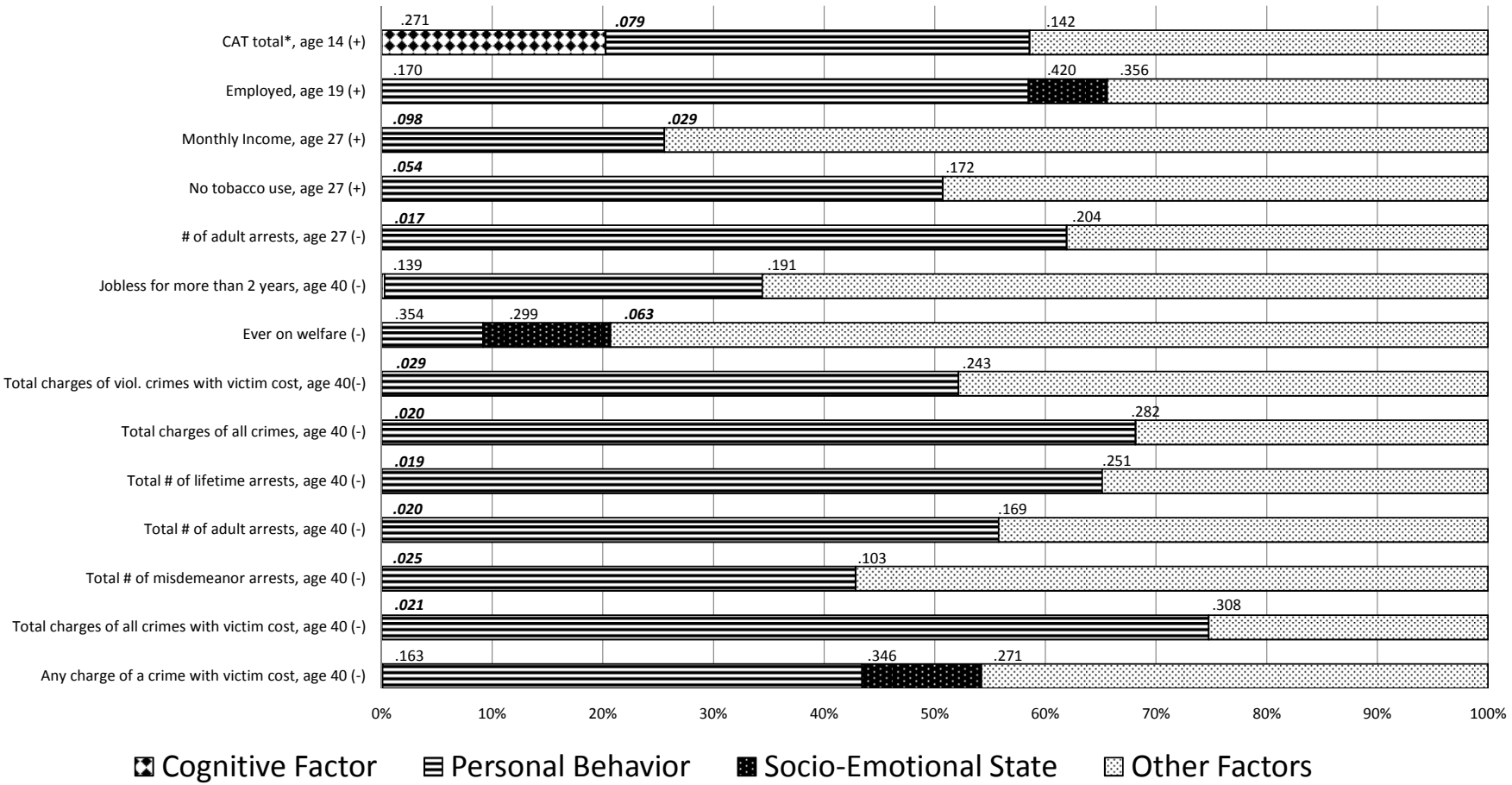


(b) Females

- Borderline statistically significant effects for both genders
- Elimination of low traits for females (again)

2. Decomposing Treatment Effects

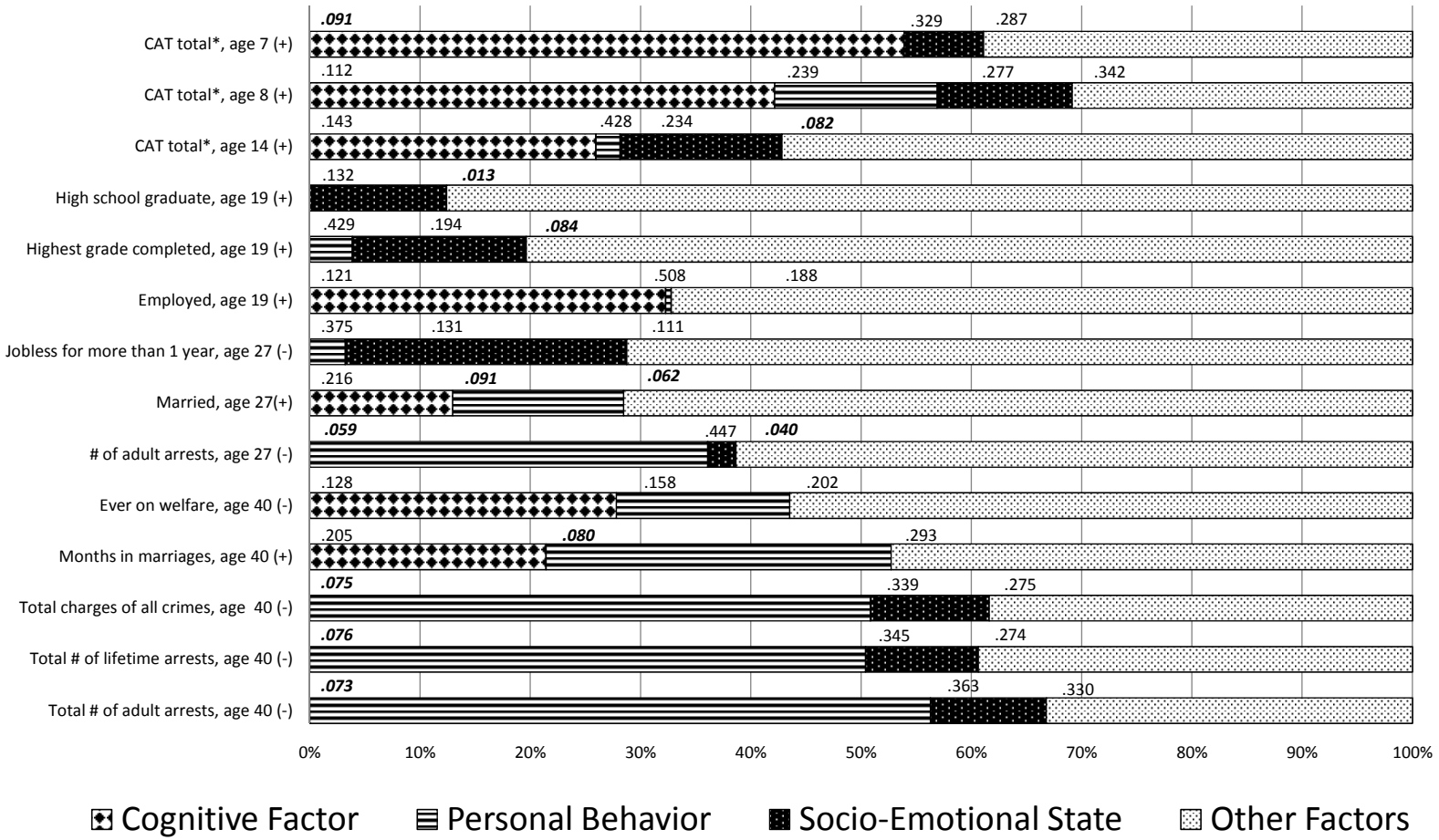
Figure 8: Decomposition of Treatment Effects, Males



Notes:

- Each bar is a decomposition of the total treatment effect into components corresponding to changes in each trait.
- For males, the intervention operates primarily through its effect on Personal Behavior.
- Contributions of Personal Behavior are generally statistically significant or borderline statistically significant, and explain up to 74% of the total treatment effects on lifetime outcomes.
- Cognition and social development play no significant role.
- Most of the statistically significant lifetime treatment effects for males are related to crime.
- Since changes in Personal Behavior are strongly predictive of changes in crime, our findings explain why crime reduction is a major benefit of the program for males (Belfield et al., 2006; Heckman et al., 2010a).
- The unexplained part is noncognitive too, since we already control for cognition.

Figure 9: Decomposition of Treatment Effects, Females



Notes: While changes in only one psychological trait explain the treatment effects for males, changes in all three traits explain the treatment effects for females, with their roles differing by outcome.
(1) Cognition

- 26–53% of the achievement test treatment effect is explained by enhancements in Cognition. The effect is statistically significant at age 7 and borderline statistically significant at ages 8 and 14.
- Effects of Cognition on employment at age 19 and on "ever being on welfare" are borderline statistically significant.

Notes (Continued):

(2) Personal Behavior

- Experimentally induced changes in Personal Behavior are important determinants of the treatment effects of marriage and crime outcomes.

(3) Socio-Emotional State

- Socio-Emotional State has borderline statistically significant effects on "high school graduation" and on "being jobless."
- The effect of Socio-Emotional State on education and crime outcomes is not precisely determined.

2. Decompositions: Gender Difference

It is instructive to distinguish two types of gender differences in the estimates from the Perry study.
(1) The sets of outcomes that exhibit at least borderline statistically significant treatment effects differ greatly across genders (see also Heckman, Moon, Pinto, Savelyev, and Yavitz, 2010b).

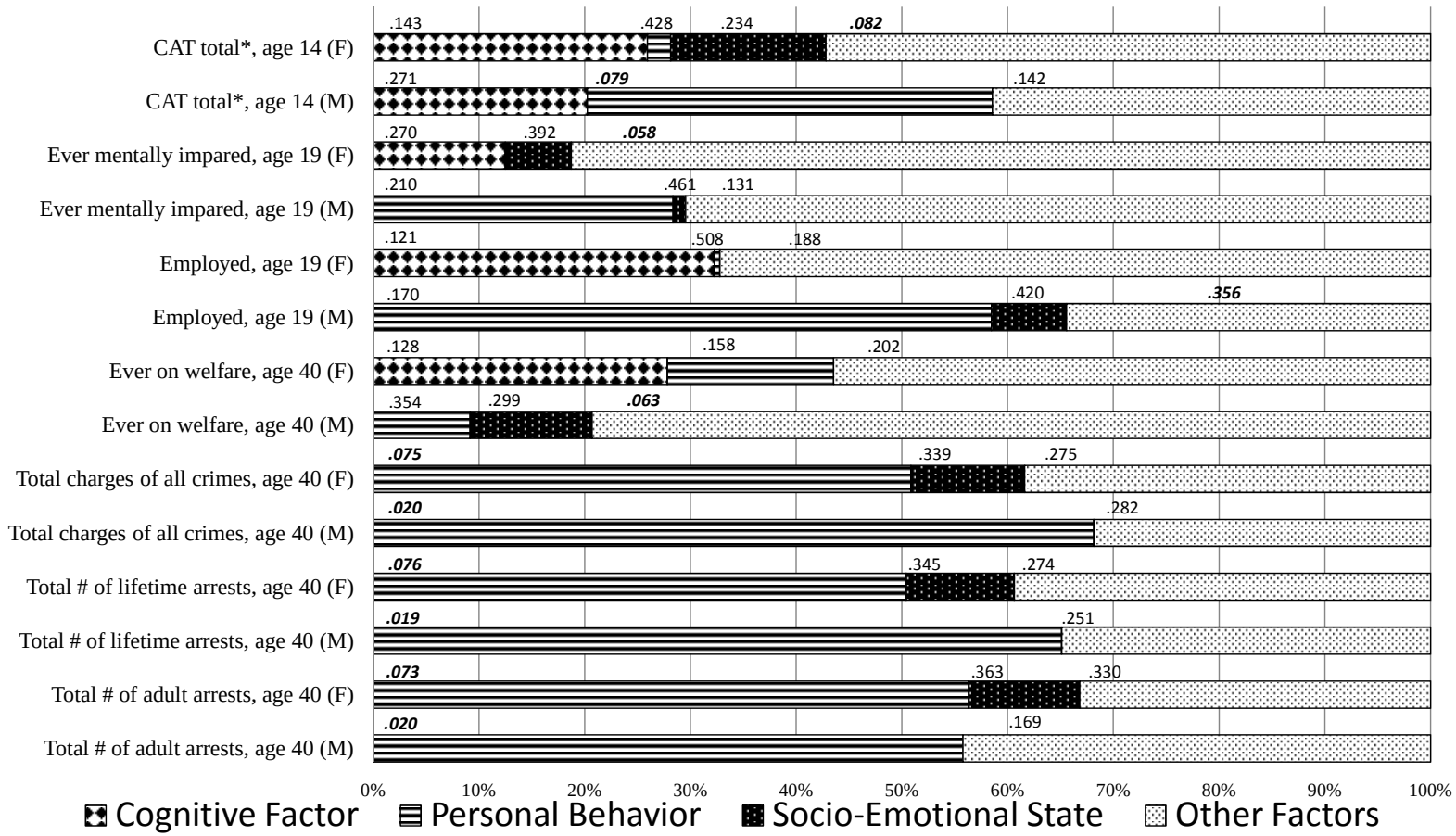
- In the figures that we have just seen different sets of outcomes are decomposed by source, and different psychological traits dominate the decompositions.
- The source of this type of gender difference is the different malleabilities of psychological traits for males and females.
- For females, all three traits are affected, which results in a wide variety of affected outcomes. For males, only Personal Behavior is affected, resulting in a smaller variety of affected outcomes.

2. Decompositions: Gender Difference (Continued)

(2) Even for outcomes showing at least borderline statistically significant treatment effects for both genders, the effects due to cognitive and noncognitive components differ.

- The second type of difference can be observed in a figure that I will show you now, comparing decompositions by gender for the same set of outcomes.
- The figure shows that differences in treatment effects between genders also arise from gender differences in the role of Cognition and Socio-Emotional State, which explain some of the experimental outcomes for females, but none for males.
- Changes in Personal Behavior affect the criminal activity of males and females in a comparable fashion.

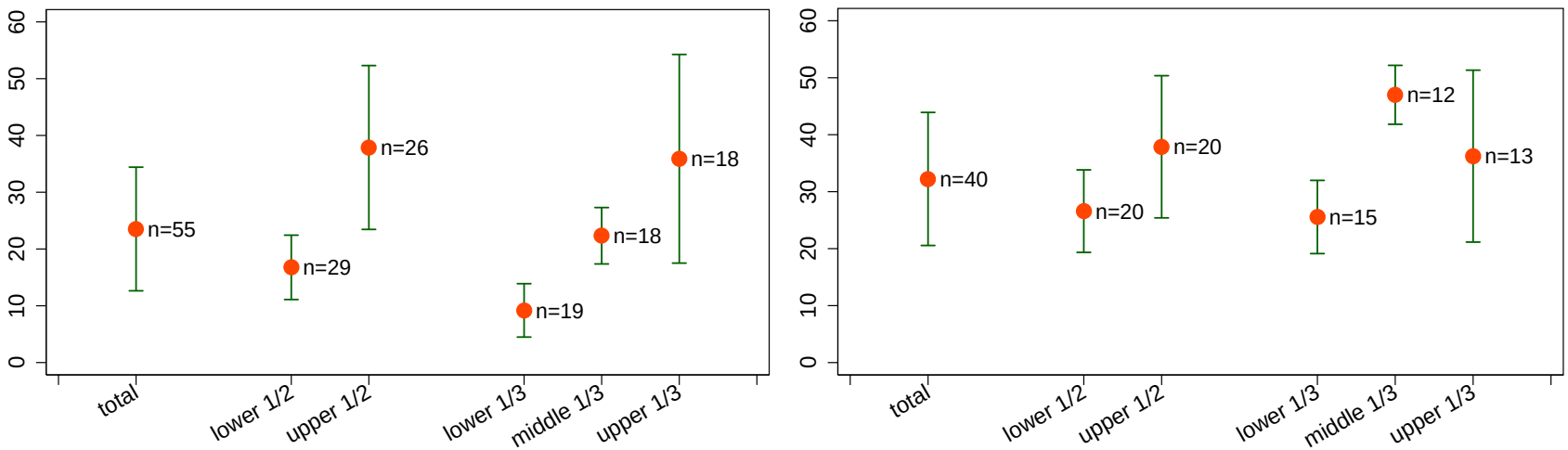
Figure 10: Decomposition of Treatment Effects, Gender Comparison



3. Quantile Treatment Effects

- Quantile treatment effects assume preservation of ranks across potential outcome distributions.
- Our analysis of the quantile treatment effects of the program shows that the program has generally stronger effects for those with higher achievement and wages, and for those more involved in crime.

Figure 11: CAT Total by Gender, age 14



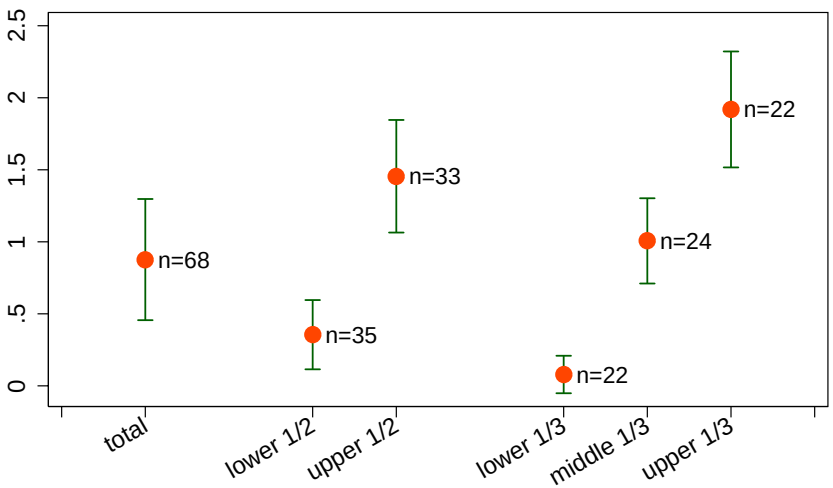
(a) Males (b) Females

- Males: effects stronger for people with higher performance
- Females: differences across quantiles weaker if present at all

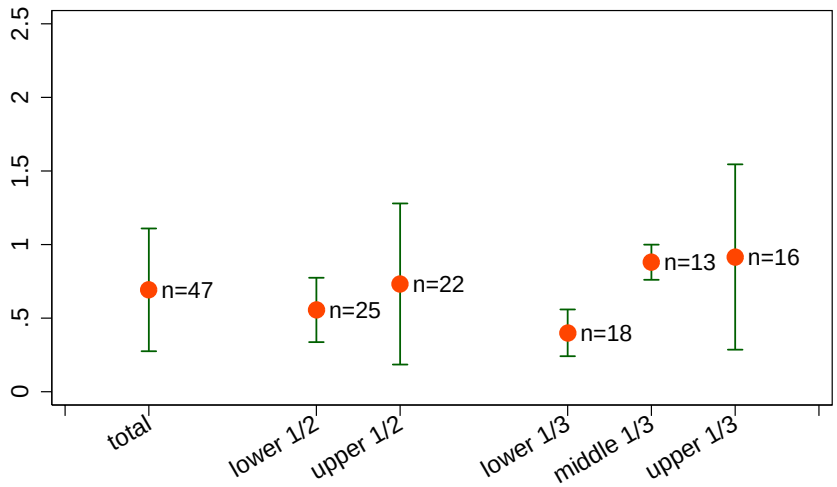
Notes:

- The figure shows unconditional differences in means for the whole sample (denoted by "total"), the 2-quantiles (denoted "lower 1/2" and "upper 1/2"), and the terciles (denoted "lower 1/3", "middle 1/3" and "upper 1/3").
- For interpretative convenience, in all subsequent figures, the higher quantile is defined to correspond to a more socially-desirable performance (higher wage, lower crime, and so on).

Figure 12: Monthly Income by Gender, age 27

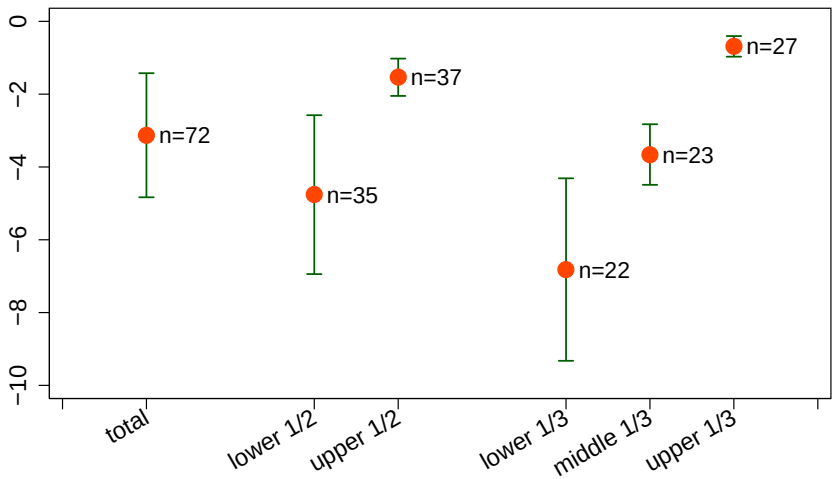


(a) Males

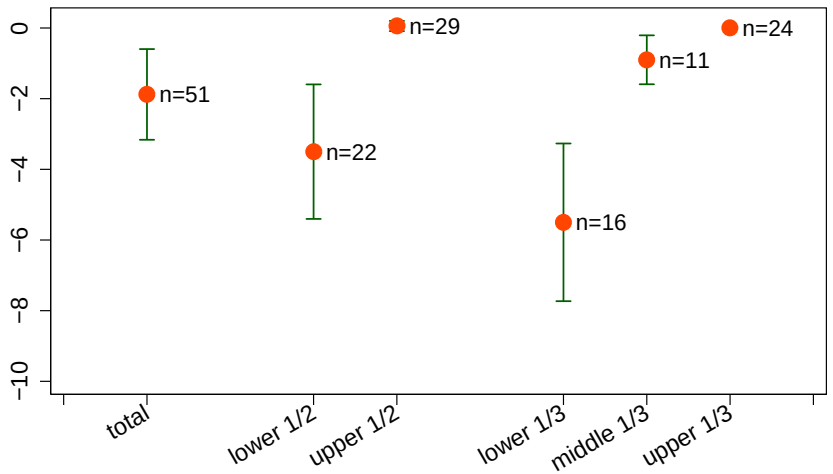


(b) Females

Figure 13: Total # of Misdemeanor Arrests by Gender, age 40



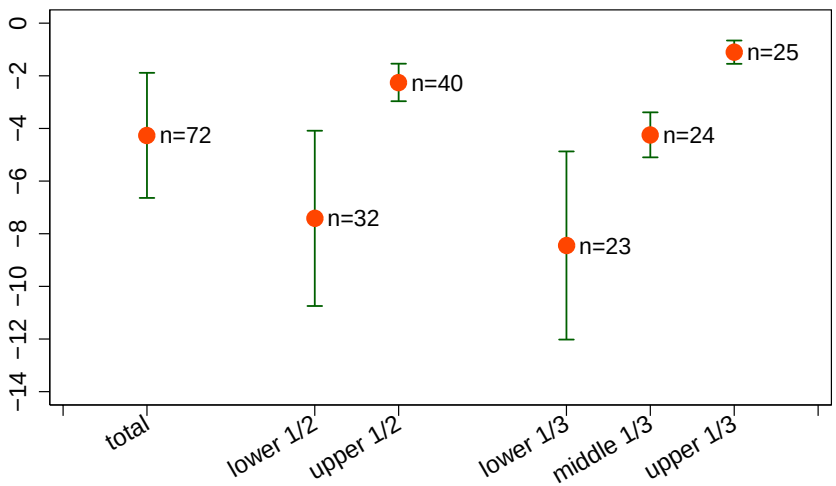
(a) Males



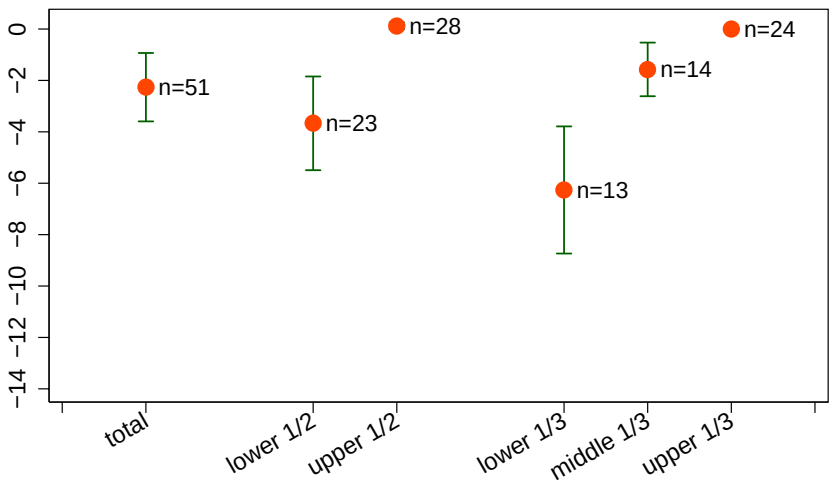
(b) Females

- Those who commit more crimes are more affected for both genders.

Figure 14: Total # of Adult Arrests by Gender, age 40

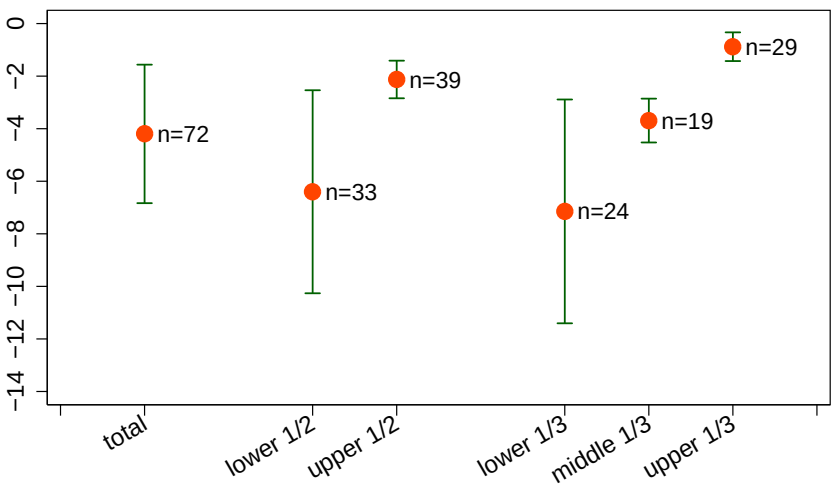


(a) Males

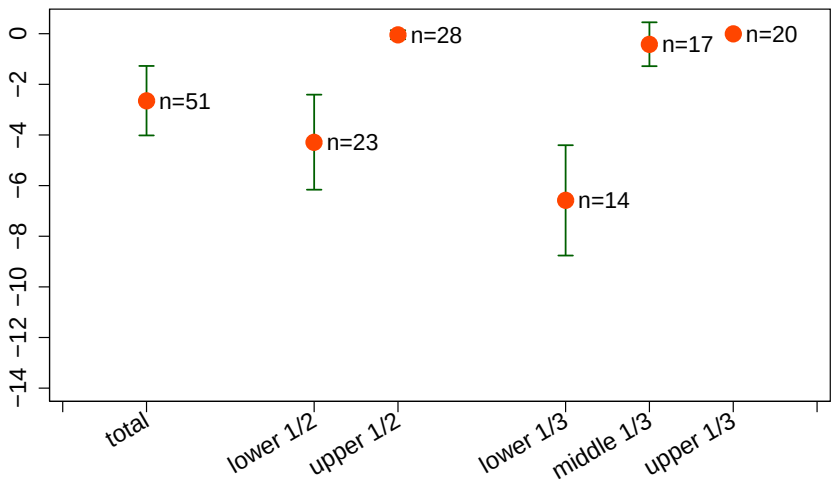


(b) Females

Figure 15: Total # of Lifetime Arrests by Gender, age 40

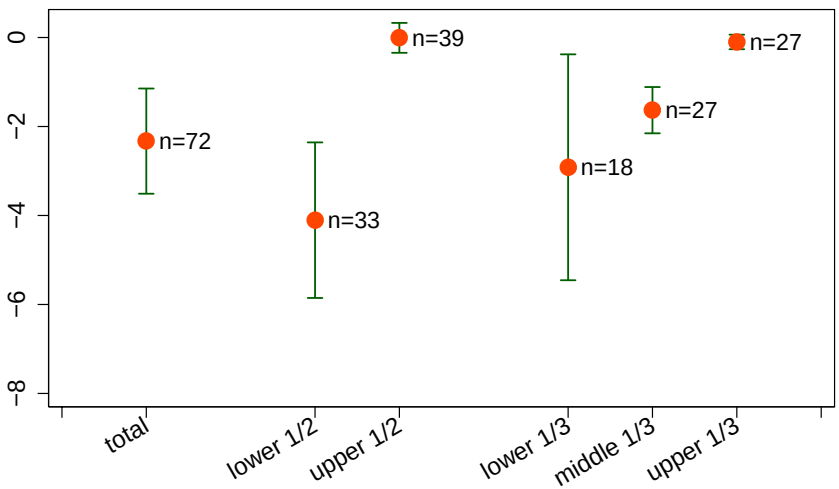


(a) Males

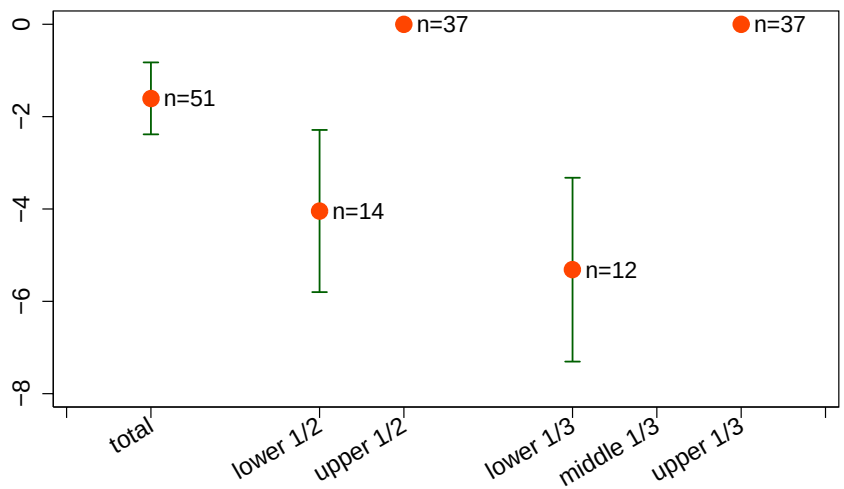


(b) Females

Figure 16: Total # of Adult Arrests by Gender, age 27



(a) Males



(b) Females

Relationship of This Work to the Literature

- This paper contributes to an emerging literature in economics documenting the predictive power of noncognitive traits (see Bowles and Gintis, 1976, 2001; Borghans, Duckworth, Heckman, and ter Weel, 2008; Heckman, Stixrud, and Urzua, 2006; Segal, 2008, 2009).
- We make five contributions.
  1. Use experimental data to identify causal effects of interventions on cognitive and noncognitive traits
  2. Use cleaner measures of cognition (IQ rather than achievement tests)
  3. Use early measures of noncognitive traits
  4. Perform decompositions of treatment effects into components due to increments in cognitive and noncognitive traits
  5. Develop a simple regression-based approach instead of more complicated maximum-likelihood methods

Conclusions

- Our analysis reveals the multiplicity of traits that produce human achievement.
- Equating human capital with one characteristic (like education level or IQ) as social scientists often do is a strong assumption that may lead to wrong conclusions.
- Noncognitive capabilities are enhanced by the intervention and play a substantial role in generating its successful outcomes.
- The importance and malleability of these traits deserve much greater emphasis in public policy discussions.