Preferences, Abilities, Personality Psychology, and Economics: Some New Results
Part II

James J. Heckman
University of Chicago

CeMMAP Masterclass
University College of London
June 23, 2016
Introduction
• Beware of pop psychology: e.g. “grit” and related claims about human performance.
• Psychologists know a lot less than they claim.
• Economics has a lot to contribute to psychology.
• Sadly, many economists are now uncritically taking measures and “insights” from psychology.
• Division of labor does not always produce greater social output.
Draw from Tim Kautz, James J. Heckman, Ron Diris, Baster Weel, and Lex Borghans in Fostering and Measuring Skills: Improving Cognitive and Noncognitive Skills to Promote Lifetime Success
OECD, 2014

Link to paper: http://tinyurl.com/lqnvb6w
Main Points

• Many important life skills are not captured by scores on achievement tests
• A sole focus on IQ or achievement test scores (e.g., NCLB, PISA and Iowa tests) gives an incomplete picture of what schools, families, and communities do and how to evaluate schools and other life cycle skill interventions
• Socioemotional skills—character, etc.—are important
• So is health
• These skills can be measured
• They are malleable, and there are effective interventions to promote them
• Soft skills more malleable than cognitive skills at later ages
• Health malleable throughout life
• Older ("established") measurement systems such as the Big Five do **not** capture the rich range of behaviors and traits that children and adults exhibit
• Economists have recently rushed to embrace them at a time when psychologists have come to question them
• Need more comprehensive measures of traits
• Economics has the tools to do this
• Ironically, they are being embraced by psychologists while other economists are embracing naive psychological measures
• Instead of relying exclusively or mainly on self-reported “Big Five measures,” there are better approaches based on behaviors
  • Teacher reports and assessments as encoded in school system records and interviews
  • Past behaviors and choices
  • Eliciting preference parameters from observed choices in the field and in controlled choice experiments and behaviors:
    • Risk aversion
    • Time preference
    • Ambiguity aversion
    • Trust
    • Reciprocity (positive and negative)
• For all measurement systems, we should adjust for incentives and other traits, something largely ignored in psychology
• There are precedents to this approach
“To value schools, by length instead of quality, is a matchless absurdity. Arithmetic, grammar, and the other rudiments, as they are called, comprise but a small part of the teachings in a school. The rudiments of feeling are taught not less than the rudiments of thinking. The sentiments and passions get more lessons than the intellect. Though their open recitations may be less, their secret rehearsals are more.”

—Horace Mann (1838)
Cognitive Skills
Modern History of Testing

- Starts with IQ tests
Figure 1: An Hierarchical Scheme of General Intelligence and Its Components

Source: Recreated from Ackerman and Heggestad [1997], based on Carroll [1993].
Cognition: “g”

• “g” is a product of early Twentieth Century psychology.
• Concept of “g” has been broadened even beyond subcomponents of “fluid” and “crystallized” intelligence.
• But still is at the center of a hierarchy of correlated traits.
• Circularity: Validation in psychology is often done using grades and other test scores.
• Rarely look at workplace or real behavioral productivity of these traits.
• Exceptions
  a. Personnel psychology
  b. AFQT and studies of achievement tests in economics
General Knowledge: The Achievement Test

- What schools add to the skills of students to perform tasks
- Assumed tests could capture performance on “real world” tasks
“We lean heavily on written examinations, on a few types of objective tests, and on the subjective impressions of teachers. Many other appraisal devices could be used, such as records of activities in which pupils participate, questionnaires, check lists, anecdotal records and observational records, interviews, reports made by parents, products made by the pupils, and records made by instruments (motion pictures, eye-movement records, sound recordings, and the like).”

—Ralph Tyler (1940)
Traditional Approach to Measurement of Personality: 
*The Big Five*
Personality Traits

- Early pioneers used a lexical approach to define personality.
- Classify words that are used to describe people.
- Culminated in the “Big Five” based on factor analysis of measurements of personality.
- Extracted from a variety of measures –
  - Observer reports
  - Tests
  - Measured productivity on the job (much less common)
- No single “g_p” explains all traits.
- Key idea: Correlations within clusters but not across clusters.
Non-Cognitive Skills
## Table 1: The Big Five Traits

### OCEAN

<table>
<thead>
<tr>
<th>Trait</th>
<th>Definition of Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Openness to Experience</td>
<td>The tendency to be open to new aesthetic, cultural, or intellectual experiences.</td>
</tr>
<tr>
<td>II. Conscientiousness</td>
<td>The tendency to be organized, responsible, and hardworking.</td>
</tr>
<tr>
<td>III. Extraversion</td>
<td>An orientation of one’s interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.</td>
</tr>
<tr>
<td>IV. Agreeableness</td>
<td>The tendency to act in a cooperative, unselfish manner.</td>
</tr>
<tr>
<td>V. Neuroticism</td>
<td>Neuroticism is a chronic level of emotional instability and proneness to psychological distress. Emotional stability is predictability and consistency in emotional reactions, with absence of rapid mood changes.</td>
</tr>
</tbody>
</table>
Table 2: The Big Five Domains and Their Facets

<table>
<thead>
<tr>
<th>Big Five Personality Factor</th>
<th>American Psychology Association Dictionary description</th>
<th>Facets (and correlated trait adjective)</th>
<th>Related Traits</th>
<th>Childhood Temperament Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness to Experience</td>
<td>“the tendency to be open to new aesthetic, cultural, or intellectual experiences”</td>
<td>Fantasy (imaginative) Aesthetic (artistic) Feelings (excitable) Actions (wide interests) Ideas (curious) Values (unconventional)</td>
<td>__</td>
<td>Sensory sensitivity Pleasure in low-intensity activities Curiosity</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>“the tendency to be organized, responsible, and hardworking’</td>
<td>Competence (efficient) Order (organized) Dutifulness (not careless) Achievement striving (ambitious) Self-discipline (not lazy) Deliberation (not impulsive)</td>
<td>Grit Perseverance Delay of gratification Impulse control Achievement striving Ambition Work ethic</td>
<td>Attention/(lack of) distractibility Effortful control Impulse control/delay of gratification Persistence Activity*</td>
</tr>
</tbody>
</table>
Table 2: The Big Five Domains and Their Facets

<table>
<thead>
<tr>
<th>Big Five Personality Factor</th>
<th>American Psychology Association Dictionary description</th>
<th>Facets (and correlated trait adjective)</th>
<th>Related Traits</th>
<th>Childhood Temperament Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>“an orientation of one’s interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability”</td>
<td>Warmth (friendly) Gregariousness (sociable) Assertiveness (self-confident) Activity (energetic) Excitement seeking (adventurous) Positive emotions (enthusiastic)</td>
<td></td>
<td>Surgency Social dominance Social vitality Sensation seeking Shyness* Activity* Positive emotionality Sociability/affiliation</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>“the tendency to act in a cooperative, unselfish manner”</td>
<td>Trust (forgiving) Straight-forwardness (not demanding) Altruism (warm) Compliance (not stubborn) Modesty (not show-off) Tender-mindedness (sympathetic)</td>
<td>Empathy Perspective taking Cooperation Competitiveness</td>
<td>Irritability* Aggressiveness Willfulness</td>
</tr>
</tbody>
</table>

*Note: Temperament traits may be related to two Big Five factors.

Source: Table adapted from John and Srivastava [1999].
Table 2: The Big Five Domains and Their Facets

<table>
<thead>
<tr>
<th>Big Five Personality Factor</th>
<th>American Psychology Association Dictionary description</th>
<th>Facets (and correlated trait adjective)</th>
<th>Related Traits</th>
<th>Childhood Temperament Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism/Emotional Stability</td>
<td>Emotional stability is “predictability and consistency in emotional reactions, with absence of rapid mood changes.” Neuroticism is “a chronic level of emotional instability and proneness to psychological distress.”</td>
<td>Anxiety (worrying) Hostility (irritable) Depression (not contented) Self-consciousness (shy) Impulsiveness (moody) Vulnerability to stress (not self-confident)</td>
<td><strong>Internal vs. External Locus of control</strong> Core self-evaluation Self-esteem Self-efficacy Optimism Axis I psychopathologies (mental disorders) including depression and anxiety disorders</td>
<td>Fearfulness/behavioral inhibition Shyness Irritability Frustration (Lack of) soothability Sadness</td>
</tr>
</tbody>
</table>

*These temperament traits may be related to two Big Five factors. Source: Table adapted from John and Srivastava [1999].

**Notes:** Facets specified by the NEO-PI-R personality inventory (Costa and McCrae [1992b]). Trait adjectives in parentheses from the Adjective Check List (Gough and Heilbrun [1983]).
• They are predictive of many outcomes, but simple correlations with meaningful outcomes are relatively low.

• The Big Five are defined without reference to any context in which they are measured (i.e., situation).

• This practice gives rise to an identification problem we discuss below.
The Person-Situation Debate

- Is variation across people in behavior a consequence of personal traits or of situations?

Mischel (*Personality and Assessment*, 1968, p. 146)

“...with the possible exception of intelligence, highly generalized behavioral consistencies have not been demonstrated, and the concept of personality traits as broad dispositions is thus untenable.”
• Many behavioral economists hold a similar view and appeal to Mischel as a guiding influence.

Thaler (2008)

“The great contribution to psychology by Walter Mischel [...] is to show that there is no such thing as a stable personality trait.”
Personality Psychology After the Person-Situation Debate

- A rich body of **correlational evidence** shows that for many outcomes, measured personality traits are as predictive, and are sometimes more predictive, than standard measures of cognition, that traits are stable across situations, but situations also matter.
- Mounting evidence that behavior has a biological basis suggests that personality is an important determinant of behavior.
- The evidence from behavioral genetics shows that measured personality traits are as heritable as cognitive traits.
- Alterations in brain structure and function through accidents, disease and by experiments affect measured personality.
Measurement Challenges
New approaches to measuring character and cognitive skills go well beyond the Big Five
• Traditional approach sanctifies the notion that a “test” captures “traits or skills”
• All psychological measurements are based on performance on some task performances
Definition of Personality

*Personality traits are the relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances.*

—Roberts (2009, p. 140)
A Task-Based Framework for Identifying and Measuring Skills

- Distinction between tasks & tests artificial
- All tests are tasks
**Figure 2: Skills, Preferences, Personality, and Performance on Tasks**

- Health Skills
- Cognitive Skills
- Personality, Character, Motivation, and Aspiration Skills
- Financial Resources
- Information
- Peers
- Environments
- Potential Actions ("Capabilities")

Actions (Functionings) Performance on Tasks

Preferences

Effort

Norms, Social Policy, Regulation
Testing the Tests
• IQ tests and achievement tests are typically validated in a circular fashion, using other measures of cognition
• Unrelated to actions in real life performance
### Table 3: Validities of GED Test

<table>
<thead>
<tr>
<th>Test</th>
<th>Correlation</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armed Forces Qualification Test (AFQT)</td>
<td>0.75 - 0.79 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>Iowa Test of Educational Development</td>
<td>0.88 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>American College Test (ACT)</td>
<td>0.80 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>Adult Performance Level (APL) Survey</td>
<td>0.81 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>New York’s Degrees of Reading Power (DRP) Test</td>
<td>0.77 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>Test of Adult Basic Education (TABE)</td>
<td>0.66-0.68 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>General Aptitude Test Battery (GATB)</td>
<td>0.61-0.67 †</td>
<td>Means and Laurence (1984)</td>
</tr>
<tr>
<td>National Adult Literacy Survey (NALS) factor</td>
<td>0.78 ‡</td>
<td>Baldwin (1995)</td>
</tr>
</tbody>
</table>

† Uses mean GED subtest scores
‡ Uses a general GED factor
### Table 4: Cognitive Ability Validities

<table>
<thead>
<tr>
<th>Test</th>
<th>Validation Domain</th>
<th>Estimate(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT (Achievement)</td>
<td>1st Year College GPA</td>
<td>0.35 - 0.53</td>
<td>Kobrin et al. (2008)</td>
</tr>
<tr>
<td>ACT (Achievement)</td>
<td>Early College GPA</td>
<td>0.42</td>
<td>ACT, Inc. (2007)</td>
</tr>
<tr>
<td>GED (Achievement)</td>
<td>HS Senior GPA</td>
<td>0.33 - 0.49</td>
<td>GED Testing Service (2009)</td>
</tr>
<tr>
<td>DAT (Achievement)</td>
<td>College GPA</td>
<td>0.13 - 0.62†</td>
<td>Omizo (1980)</td>
</tr>
<tr>
<td>AFQT (Achievement)</td>
<td>9th Grade GPA</td>
<td>0.54</td>
<td>Borghans et al. (2011)</td>
</tr>
<tr>
<td>WAIS (IQ)</td>
<td>College GPA</td>
<td>0.38 - 0.43</td>
<td>Feingold (1982)</td>
</tr>
<tr>
<td>WAIS (IQ)</td>
<td>HS GPA</td>
<td>0.62</td>
<td>Feingold (1982)</td>
</tr>
<tr>
<td>Various IQ**</td>
<td>9th Grade GPA</td>
<td>0.42</td>
<td>Borghans et al. (2011)</td>
</tr>
<tr>
<td>WISC (IQ)</td>
<td>WRAT (Achievement)</td>
<td>0.44 - 0.75‡</td>
<td>Hartlage and Steele (1977)</td>
</tr>
</tbody>
</table>
# Table 4: Cognitive Ability Validities

<table>
<thead>
<tr>
<th>Test</th>
<th>Validation Domain</th>
<th>Estimate(s)</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WISC-R (IQ)</td>
<td>WRAT (Achievement)</td>
<td>0.35 - 0.76‡</td>
<td>Hartlage and Steele (1977)</td>
</tr>
<tr>
<td>Various IQ**</td>
<td>AFQT (Achievement)</td>
<td>0.65</td>
<td>Borghans et al. (2011)</td>
</tr>
<tr>
<td>Stanford Binet (IQ)</td>
<td>WISC-R (IQ)</td>
<td>0.77 - 0.87</td>
<td>Rothlisberg (1987), Greene et al. (1990)</td>
</tr>
<tr>
<td>Raven’s (IQ)</td>
<td>WAIS-R (IQ)</td>
<td>0.74 - 0.84</td>
<td>O’Leary et al. (1991)</td>
</tr>
<tr>
<td>WIAT (Achievement)</td>
<td>CAT/2 (Achievement)</td>
<td>0.69 - 0.83*</td>
<td>Michalko and Saklofske (1996)</td>
</tr>
</tbody>
</table>

‡ Large range is due to varying validity of eight subtests of DAT
‡‡ Ranges are given because correlations vary by academic subject
* Ranges are given because correlations vary by grade level
** IQ is pooled across several IQ tests using IQ percentiles

Table 5: Correlations Among NLSY79 Measures of Cognition

<table>
<thead>
<tr>
<th></th>
<th>Correlation between IQ, AFQT, and GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
</tr>
<tr>
<td>IQ</td>
<td>1</td>
</tr>
<tr>
<td>AFQT</td>
<td>0.65</td>
</tr>
<tr>
<td>GPA(9th)</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Source: National Longitudinal Survey of Youth (NLSY79). Pooled male and female random sample. Notes: The Armed Forces Qualifying Test (AFQT) was administered in 1980 when subjects were 15-22. AFQT is adjusted for schooling at the time of the test conditional on final schooling, following the procedure in Hansen et al. (2004). AFQT is constructed from Arithmetic Reasoning, Word Knowledge, Math Knowledge, and Paragraph Comprehension tests. IQ and GPA are from high school transcripts. IQ is pooled across several IQ tests using IQ percentiles. GPA is the individual’s core-subject GPA measured in 9th grade when virtually all sample participants are enrolled. Differences between males and females are slight. For the sake of brevity we report pooled results.
Figure 3: Validities of Cognitive Measures in Age-35 Labor Market Outcomes (Adjusted R-Squared)

(a) Males
Figure 4: Validities of Cognitive Measures in Age-35 Labor Market Outcomes (Adjusted R-Squared)

(b) Females
• Much of the variance in outcomes is not explained. Lots of room for uncovering other dimensions of human performance.
Additional Validity Studies
Main Findings from Predictive Analyses

- **Conscientiousness** is the most predictive Big Five trait across many outcomes.
  a. Educational attainment, grades
  b. Job performance across a range of occupational categories
     (predictive power of “g” decreases with job complexity)
  c. Longevity
  d. Criminality

- **Neuroticism** (and related *locus of control*)
  a. Predicts schooling outcomes
  b. Labor market search

- Other traits play roles at finer levels.
Facets related to Emotional Stability (the opposite of Neuroticism) are also important for labor market success. However, accounting for reverse causality is particularly important because strong evidence suggests that labor market participation can affect traits related to Neuroticism (see the discussion of Gottschalk, 2005, in Section 8). Several studies have addressed this problem by using measures of personality measured well before individuals enter the labor market and find that locus of control and self-esteem, two facets of Emotional Stability, predict wages (Judge and Hurst, 2007; Drago, 2008; Duncan and Dunifon, 1998). Table 1.11 presents results from the structural model of Heckman, Stixrud, and Urzua (2006), suggesting that standardized adolescent measures of locus of control and self-esteem predict adult earnings to a similar degree as cognitive ability. However, the effects vary across educational levels. In general, their measure of noncognitive ability (personality) affects wages to a similar degree across all education levels, whereas cognitive ability tends to have little effect for GED recipients, high-school dropouts, and college dropouts.

However, more recent evidence suggests that personality affects wages mostly through the channel of educational attainment. In Section 7.1, we presented evidence that personality measures (along with measurements of cognition) are strong predictors of educational attainment. Heckman, Humphries, Urzua, and Veramendi (2011) estimate...
Personality may affect health-related behavior, such as smoking, diet, and exercise. For example, Hampson, Goldberg, Vogt, and Dubanoski (2007) find that high scores of teacher assessments of Extraversion, Agreeableness, and Conscientiousness during elementary school predict overall health behaviors during midlife (less smoking, more exercise, better self-rated health) and indirectly affect health through educational attainment. The effects that were statistically significant at the 5% level or less ranged from 0.06 for the effect of Extraversion on physical activity to 0.12 for the effect of Conscientiousness on self-reported health status. Both the initial level and the growth in hostility (a facet of Neuroticism) throughout elementary school predict cigarette, alcohol, and marijuana use in high school, and sociability (a trait related to Extraversion) predicts drinking but not smoking (Hampson, Tildesley, Andrews, Luyckx, and Mroczek, 2010). As Fig. 1.19 illustrates, Heckman, Stixrud and Urzua (2006) find that their personality factor affects the probability of daily smoking for males. The gradient is steepest at the high and low quantiles of the distribution.

Source: Roberts et al. (2007).
Personality and Crime
**Figure 7: Association of the Big Five and intelligence with years of completed schooling**

*Notes:* The figure displays standardized regression coefficients from a multivariate regression of years of school attended on the Big Five and intelligence, controlling for age and age squared. The bars represent standard errors. The Big Five coefficients are corrected for attenuation bias. The Big Five were measured in 2005. Years of schooling were measured in 2008. Intelligence was measured in 2006. The measures of intelligence were based on components of the Wechsler Adult Intelligence Scale (WAIS). The data is a representative sample of German adults between the ages 21 and 94. *Source:* Almlund et al. (2011) German Socio-Economic Panel (GSOEP), waves 20042008.
Few studies have examined the relationship between the Big Five and criminal behavior. The available evidence suggests that Big Five Conscientiousness and Agreeableness are important protective factors against criminal activity. Figure 1.21 illustrates that in a sample of at-risk youth, boys who had committed severe delinquent behaviors were more than three quarters of a standard deviation lower in Agreeableness and Conscientiousness, as measured by mother’s reports at age 12 or 13, than boys who had committed minor or no delinquent behaviors up to that age (John, Caspi, Robins, and Moffitt, 1994).

Much of the literature in criminology focuses on the effects of self-control on crime. People with low self-control are “impulsive, insensitive, physical (as opposed to mental), risk taking, short sighted, and nonverbal” (Gottfredson and Hirschi, 1990, p. 90). Measures of self-control are associated with Big Five Conscientiousness (O’Gorman and Baxter, 2002). Several studies have confirmed that self-control is associated with criminal activity. In an international sample, controlling for basic demographics, a measure

Source: John et al. (1994).
What Do Achievement Tests Measure?

Figure 9: Decomposing Variance Explained for Achievement Tests and Grades into IQ and Character: Stella Maris Secondary School, Maastricht, Holland

Source: Borghans et al. (2011).

Note: Grit is a measure of persistence on tasks (Duckworth et al., 2007).
Figure 10: Predictive Power ($R^2$) of Psychological Measures for Grades and Achievement Tests: Standardized Regression Coefficients

Notes: All dependent and independent variables are standardized (mean zero, st.dev. one).
What Do Grades and Achievement Tests Measure?
Lex Borghans, Bart H. H. Golsteyn, James J. Heckman and John Eric Humphries (2015)
Figure 11: Decomposing Achievement Tests and Grades into IQ and Personality

(A) Stella Maris
Figure 11: Decomposing Achievement Tests and Grades into IQ and Personality (cont.)

(B) British Cohort Study
Figure 12: Decomposing Life Outcomes into IQ and Personality
Figure 13: Decomposing Life Outcomes into IQ and Personality (NLSY79)
Figure 14: Decomposing Life Outcomes into Cognition and Personality (MIDUS)
Evidence from the General Educational Development (GED) Program

- An informative natural experiment
Figure 14: Distribution of Cognitive Ability by Educational Status

- **HS Dropout**
- **GED, no college**
- **HS Grad., no college**
Figure 15: Distribution of Character Skills by Education Group
Figure 16: Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates: Smoking and Drinking

Figure 16: Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates: Sex and Violent Behavior

Figure 16: Measures of Adolescent Behaviors for Male Dropouts, GED Recipients, and High School Graduates: Criminal Behavior

**Figure 17:** Hourly Wage Differences of GED Recipients and Traditional Graduates Compared to Uncertified Dropouts—Ages 20–39

(a) Male

[Graph showing hourly wage differences]
**Figure 17:** Hourly Wage Differences of GED Recipients and Traditional Graduates Compared to Uncertified Dropouts—Ages 20–39
Link to Appendix
Economic Models of Personality and Its Implications for Measurement of Personality and Preference
How to Conceptualize These Correlations and Establish a Causal Basis for Them?

a. Place the concept of personality within economic model(s).

b. *Personality as a strategy*: Define personality as an emergent property of a system.

c. Use the economic model(s) to frame and solve a central identification problem in empirical psychology (cognitive and noncognitive).

d. How to go from measurements of personality to personality traits.
Economic Frameworks for Conceptualizing and Measuring Personality and Personality Traits
How to interpret personality measurements within economic models?

Through

- Preferences? (Standard Approach) – But which preferences?
- Constraints? (Borghans, Duckworth, Heckman and ter Weel, 2008) or
- Expectations? (Several Recent Papers) or
- Strategies? (Social Interaction and Situation)
- All four
All measures are captured by performance on tasks

- All measurement systems in psychology are based on performance on these tasks gauged in various ways.
- Taking an IQ test is a task.
- Reporting a personality trait is a task.
- Distinction between traits and tasks is flimsy.
Tasks

- Output on task $j$, $\phi_j (\theta, e_j)$;
- $e_j$ is effort in $j$;
- $\phi_j(\cdot)$ concave and increasing in $e_j$.
- $P_j = \phi_j(\cdot)$.
- $R_j$ is reward.
- Agent maximizes

$$\sum_{j=1}^{J} R_j \phi_j (\theta, e_j)$$

subject to

$$\sum_{j=1}^{J} e_j = \bar{e}.$$

- $R_j \uparrow e_j \uparrow$
- Agent might specialize if there are increasing returns.
- Could add a cost of effort and differential costs of effort – by category.
• Preferences capture “goals.”

• A utility function over $C$ and $e$ with preference parameter vector $\psi \in \Psi$.

• $C = \Sigma R_j \phi_j(\theta, e_j)$.

• $U(C, e, \psi | I')$

• where $e = (e_1, \ldots, e_J)$.

• $I$: information (may depend on $\theta$).
- Captures the notions that
  (a) Agents have preferences over consumption $C$ and effort
  (b) Agents may value the output of tasks in their own right
  (c) Traits may facilitate dealing with uncertainty ($I$ depends on $\theta$)
An Economic Definition of Personality

- **Personality traits** are components of $e, \theta$.
- We observe **measured personality**—behaviors generated by incentives, goals, and traits.
• One might define measured personality as the performance (the $P_j$) and effort (the $e_j$) that arise from solutions to the optimization problems previously stated.

• This approach does not capture the full range of behaviors considered by personality psychologists that constitute aspects of personality.

• The actions considered by psychologists include a variety of activities that economists normally do not study, e.g., cajoling, beguiling, bewitching, charming, etc.

• To capture these more general notions, we introduce a set of “actions” broader than what is captured by $e$. 
Actions or Strategies Adopted

- Actions are styles of behavior that affect how tasks are accomplished.
- They include aspects of behavior that go beyond effort as we have defined it.
- Tasks can be accomplished by taking actions.
- The $i^{th}$ possible action to perform task $j$: $a_{i,j}$, $i \in \{1, \ldots, K_j\}$.
- Array actions in a vector $a_j = (a_{1,j}, \ldots, a_{K_j,j}) \in A$.
- The actions may be the same or different across the tasks.
- The actions are strategies agents take in response to situations.
• The productivity of the agent in task $j$ depends on the actions taken in that task:

$$P_j = \tau_j \left( a_{1,j}, a_{2,j}, \ldots, a_{K_j,j} \right).$$

(2)

• The actions themselves depend on traits $\theta$ and “effort” $e_{i,j}$:

$$a_{i,j} = \nu_{i,j} (\theta, e_{i,j})$$

(3)

where

$$\sum_{i=1}^{K_j} e_{i,j} = e_j \text{ and } \sum_{j=1}^{J} e_j = \bar{e}.$$ 

• Actions generalize the notion of effort to a broader class of behavior.
Agents may have utility over actions beyond the utility they get from consuming the outputs of tasks.

Define utility over actions.

Let $a$ denote the choice of actions applied to all tasks: $(a = (a_1, \ldots, a_J))$.

$\mathcal{M}$: the set of actions, including actions that do not directly contribute to productivity.

$$a_{i,m} = \nu_{i,m} (\theta, e_{i,m}), \ m \in \mathcal{M}$$

$\mathcal{A} \subseteq \mathcal{M}$. 
The agent solves

$$\max E \left[ U(a, C, P, e | \psi) | I \right]$$

with respect to $e$ given the stated constraints.
Introducing Situations

- Situations indexed by $h \in \mathcal{H}$.
- For a person with traits $\theta$ and effort vector $e_j$ with action $a_{i,j}$, using the specification (3), the action function can be expanded to be dependent on situation $h$:

$$a_{i,j,h} = \nu_{i,j}(\theta, e_{i,j,h}, h),$$

(productivity on a task)

$$P_{j,h} = \tau_j(a_{1,j,h}, \ldots, a_{K,j,h})$$

or more generally

$$P_{j,h} = \tau_j(\theta, a_{1,j,h}, \ldots, a_{K,j,h}, h).$$
• Equations (4)–(6) resolve the person-situation debate.
• Failure to control for situation $h$, just like failure to control for effort, contaminates identification of traits using measures of actions or productivities.
• Let $T \in T$ be the vector of traits $(\theta, \psi, \bar{e})$.
• The solution to the general constrained maximization problem is to pick goods $X$, situation $h$, action $a_{i,j}$, and effort $e_j$, $j \in \{1, \ldots, J\}$ subject to the constraints.
• $h$ is fixed if the situation is forced on the agent.
• For simplicity, we analyze this case.
• More generally, situations chosen.
• The situations are (strategic) interactions among agents.
• People may have different personalities depending on their trait endowments, constraints, and situations.

• The actions – not the traits – constitute the data used to identify the traits.

• Personality psychologists use actions (e.g., “dispositions”) to infer traits.

• The same identification issues previously discussed apply to a broader set of measurements of behaviors.
Many personality psychologists define personality as

“enduring patterns of thoughts, feelings and behaviors”

that reflect tendencies of persons to respond in certain ways under certain circumstances.
• What are enduring patterns of actions?

• “Enduring actions”—average of the \( a \) functions for a person with a given trait vector \( T = t \) over situations and efforts: 
  \[
  T = (\theta, \psi, \bar{e})
  \]
• For task $j$ and trait vector $t$, the average action for information set $\mathcal{I}$ can be defined as

$$\bar{a}_{T,j,\mathcal{I}} = \int_{S_{T,\mathcal{I}}(h,e_{i,j})} \nu_{i,j} (\theta, e_{i,j}, h) \ g \ (h, e_{i,j} \mid T = (\theta, \psi, \bar{e}), \mathcal{I}) \ dh \ de_{i,j}. $$

• $S_{T,\mathcal{I}}(h, e_{i,j})$ is the support of $(h, e_{i,j})$ given $T$ and $\mathcal{I}$. 
• $g(h, e_{i,j} \mid T = (\theta, \psi, \bar{e}), \mathcal{I})$ is the density of $(h, e_{i,j})$ given $T = (\theta, \psi, \bar{e})$ and information set $\mathcal{I}$.

• $\bar{a}_{T,j,\mathcal{I}}$ is the “enduring action” of agents across situations in task $j$ with information $\mathcal{I}$, i.e., the average personality.

• If $\nu_{i,j}$ is separable in $T$, the marginal effect of personality trait vector $\theta$ is the same in all situations.
• One can define the “enduring traits” in a variety of ways, say by averaging over tasks, $j$, situations, $h$, or both.

• Only under separability in $T$ will one obtain the same marginal effect of $\theta$.

• Epstein (1979) and a subsequent literature present evidence in favor of an “enduring trait” that is common across situations.
Psychological Variables as Constraints: Another Way to Conceptualize Personality

Link to Appendix
A Third Model

Incorporating Personality and Cognitive Ability into Conventional Economic Models: A Simple Framework for Organizing the Evidence

Link to Appendix
Personality and Preference Parameters
• Measures of personality predict a wide range of life outcomes.
• Personality psychologists define traits as relatively stable, person-specific determinants of behavior: preferences are the natural counterpart of these traits in economics.
• Preferences are also, at least in many models, unaffected by changes in constraints.
• While personality might relate to preferences, the exact link is unclear.
Table 6: Standard Preference Parameters and Conceptually Similar Measures in the Psychology Literature

<table>
<thead>
<tr>
<th>Preference parameter</th>
<th>Personality measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time preference</td>
<td>Conscientiousness</td>
</tr>
<tr>
<td></td>
<td>Self-control</td>
</tr>
<tr>
<td></td>
<td>Affective mindfulness</td>
</tr>
<tr>
<td></td>
<td>Consideration of future consequences</td>
</tr>
<tr>
<td></td>
<td>Elaboration of consequences</td>
</tr>
<tr>
<td></td>
<td>Time preference</td>
</tr>
<tr>
<td>Risk aversion</td>
<td>Impulsive sensation seeking</td>
</tr>
<tr>
<td></td>
<td>Balloon Analogue Risk Task</td>
</tr>
<tr>
<td>Leisure Preference</td>
<td>Achievement Striving</td>
</tr>
<tr>
<td></td>
<td>Endurance</td>
</tr>
<tr>
<td></td>
<td>Industriousness</td>
</tr>
<tr>
<td>Social preference</td>
<td>Warmth</td>
</tr>
<tr>
<td></td>
<td>Gregariousness</td>
</tr>
<tr>
<td></td>
<td>Trust</td>
</tr>
<tr>
<td></td>
<td>Altruism</td>
</tr>
<tr>
<td></td>
<td>Tender-mindedness</td>
</tr>
<tr>
<td></td>
<td>Hostility</td>
</tr>
</tbody>
</table>
Table 7: Overview of Empirical Studies of the Links Between Preferences and Traits

<table>
<thead>
<tr>
<th>Preferences</th>
<th>Personality measure</th>
<th>Empirical study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Preference</td>
<td>Conscientiousness, Self-control, Affective mindfulness, Elaboration of consequences, Consideration of future consequences.</td>
<td>Daly, Delaney and Harmon [2009]</td>
</tr>
<tr>
<td></td>
<td>Extraversion</td>
<td>Dohmen, Falk, Huffman et al. [2010]</td>
</tr>
<tr>
<td></td>
<td>Time Preference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Openness</td>
<td>Dohmen, Falk, Huffman et al. [2010]</td>
</tr>
<tr>
<td></td>
<td>Neuroticism, ambition, Agreeableness</td>
<td>Borghans, Golsteyn, Heckman et al. [2009]</td>
</tr>
<tr>
<td></td>
<td>Balloon Analogue Risk Task</td>
<td>Lejuez, Aklin, Zvolensky et al. [2003]</td>
</tr>
<tr>
<td>Social Preferences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altruism</td>
<td>Neuroticism, Agreeableness</td>
<td>Ashton, Paunonen, Helmes et al. [1998], Osiński [2009], Bekkers [2006]</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>Neuroticism, Agreeableness, Conscientiousness</td>
<td>Dohmen, Falk, Huffman et al. [2008]</td>
</tr>
<tr>
<td>Trust</td>
<td>Neuroticism, Agreeableness, Openness, Conscientiousness</td>
<td>Dohmen, Falk, Huffman et al. [2008]</td>
</tr>
</tbody>
</table>

See ADHK for more complete discussion.
Investigating the Link: The Relationship Between Economic Preferences and Psychological Personality Measures
Table 8: Overview of the experimental measures in data set from laboratory experiments among university students

<table>
<thead>
<tr>
<th>Preference</th>
<th>Experiment</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Two lists of choices between an amount of money “today” and an amount of money “in 12 months”.</td>
<td>Average switching point over both lists of choices from the early to the delayed amount.</td>
</tr>
<tr>
<td>Risk</td>
<td>Two lists of choices between a lottery and varying safe options.</td>
<td>Average switching point over both lists of choices from the lottery to the safe option.</td>
</tr>
<tr>
<td>Positive Reciprocity</td>
<td>Second-mover behavior in two versions of the trust game (strategy method).</td>
<td>Average amount sent back in both trust games.</td>
</tr>
<tr>
<td>Negative Reciprocity</td>
<td>Investment into punishment after unilateral defection of the opponent in a prisoner’s dilemma (strategy method).</td>
<td>Amount invested into punishment.</td>
</tr>
<tr>
<td>Trust</td>
<td>First mover behavior in two versions of the trust game.</td>
<td>Average amount sent as a first mover in both trust games.</td>
</tr>
<tr>
<td>Altruism</td>
<td>First mover behavior in a dictator game with a charitable organization as recipient.</td>
<td>Size of donation.</td>
</tr>
</tbody>
</table>
Correlation Structure
Table 9: Experimental Data: Pearson correlation structure experimental data set

<table>
<thead>
<tr>
<th></th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Neuroticism</th>
<th>LoC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.0370</td>
<td>0.0057</td>
<td>−0.0084</td>
<td>0.1026**</td>
<td>−0.0518</td>
<td>0.0847</td>
</tr>
<tr>
<td>Risk</td>
<td>−0.0379</td>
<td>−0.0611</td>
<td>0.0762*</td>
<td>0.0202</td>
<td>−0.1201***</td>
<td>0.0434</td>
</tr>
<tr>
<td>Pos. Reciprocity</td>
<td>0.1724***</td>
<td>0.0140</td>
<td>0.0211</td>
<td>0.2042***</td>
<td>0.0361</td>
<td>0.0152</td>
</tr>
<tr>
<td>Neg. Reciprocity</td>
<td>−0.0885*</td>
<td>−0.0393</td>
<td>0.0943*</td>
<td>−0.1451***</td>
<td>−0.0136</td>
<td>−0.1418**</td>
</tr>
<tr>
<td>Trust</td>
<td>0.1232***</td>
<td>−0.1300***</td>
<td>0.0004</td>
<td>0.1665***</td>
<td>−0.0134</td>
<td>−0.0140</td>
</tr>
<tr>
<td>Altruism</td>
<td>0.1242**</td>
<td>−0.0979*</td>
<td>0.0249</td>
<td>0.1911***</td>
<td>0.0847*</td>
<td>0.0480</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Correlations between economic preferences and the Big Five were calculated using 394–477 observations. Correlations between economic preferences and locus of control were calculated using between 254–315 observations. All measures are standardized.
### Table 10: Representative Panel Data: Pearson correlation structure between personality measures and economic preferences from SOEP observations

<table>
<thead>
<tr>
<th></th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Neuroticism</th>
<th>LoC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.0183**</td>
<td>0.1122***</td>
<td>-0.0415***</td>
<td>0.3122***</td>
<td>-0.0584***</td>
<td>0.0681***</td>
</tr>
<tr>
<td>Risk</td>
<td>0.2793***</td>
<td>-0.0400***</td>
<td>0.2601***</td>
<td>-0.1454***</td>
<td>-0.0996***</td>
<td>0.1521***</td>
</tr>
<tr>
<td>Pos. Reciprocity</td>
<td>0.1814***</td>
<td>0.2520***</td>
<td>0.1473***</td>
<td>0.1842***</td>
<td>0.0872***</td>
<td>0.0954***</td>
</tr>
<tr>
<td>Neg. Reciprocity</td>
<td>-0.0522***</td>
<td>-0.1558***</td>
<td>-0.0264***</td>
<td>-0.3756***</td>
<td>0.0612***</td>
<td>-0.2154***</td>
</tr>
<tr>
<td>Trust</td>
<td>0.1272***</td>
<td>-0.0680***</td>
<td>0.0575***</td>
<td>0.0945***</td>
<td>-0.1919***</td>
<td>0.2094***</td>
</tr>
<tr>
<td>Altruism</td>
<td>0.1756***</td>
<td>0.1495***</td>
<td>0.1670***</td>
<td>0.2557***</td>
<td>0.0908***</td>
<td>0.0874***</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Correlations are calculated using 14,243 observations. All measures are standardized.
Adjusted $R^2$'s for linear regressions for life outcomes. The number of observations available varies for the different life outcomes: subjective health (14,218), life satisfaction (14,214), gross wage (7,199), unemployed (9,095), and years of education (13,768). Gross wage measures the gross hourly wage.
Link to Appendix for Becker et al.
Further Information on the Relationship Between Economic Preferences and Conventional Personality Measures
Other Research

Altruism and Social Preferences
Identifying Personality “Traits” from Measured Performance on Tasks

- Key assumption: Some tasks may require only a single trait or a subset of all of the traits.
• Using performance on a task (or on multiple measures of the task) to identify a “trait” requires that performance on certain tasks (performance on a test, performance in an interpersonal situation, etc.) depends exclusively on one component of $\theta$, say $\theta_{1,j}$, and we standardize for incentives and effort.
• Assumes task \( j \) output is

\[
P_j = \phi_j (\theta_{1,j}, e_j).
\]

• One must standardize for the effort at a benchmark level, say \( e^* \), to use \( P_j \) to identify a measure of the trait \( \theta_{1,j} \).
• The activity of picking a task (or a collection of tasks) that measure a particular trait ($\theta_{1,j}$ in our example) is called **operationalization** in psychology.

• Demonstrating that a measure successfully operationalizes a trait is called **construct validity**.

• *Need to standardize for effort to measure the trait.*

• Otherwise produces variation in the measured trait across situations with different incentives.
A Fundamental Identification Problem

- Operationalization and construct validation require heroic assumptions.
- Even if one adjusts for effort in a task, productivity in a task may depend on *multiple traits*.
- Thus two components of $\theta$ (say $\theta_{1,\mu}$, $\theta_{1,\pi}$) may determine productivity in $j$.
- Without further information, one cannot infer which of the two traits produces the productivity in $j$.
- In general, even having two (or more) measures of productivity that depend on $(\theta_{1,\mu}, \theta_{1,\pi})$ is not enough to identify the separate components.
• Ignore measurement error for now.
• Consider the following case of two productivity measures for the two tasks $j$ and $j'$:

\[
P_j = \phi_j\left(\theta_{1,\mu}, \theta_{1,\pi}, e_j\right)
\]
\[
P_{j'} = \phi_{j'}\left(\theta_{1,\mu}, \theta_{1,\pi}, e_{j'}\right), \quad j \neq j'.
\]

• Standardize measurements at a common level of effort $e_j = e_{j'} = e^*$.
• Note that if the support of $e_j$ and $e_{j'}$ is disjoint, no $(\theta_{1,\mu}, \theta_{1,\pi})$ uniquely defined.
• If the system of equations satisfies a local rank condition, then one can solve for the pair $(\theta_{1,\mu}, \theta_{1,\pi})$ at $e^*$. 
• Note, however, that **only the pair is identified.**

• One cannot (without further information) determine which component of the pair is \( \theta_{1,\mu} \) or \( \theta_{1,\pi} \).

• In the absence of **dedicated constructs** (constructs that are generated by only one component of \( \theta \)), there is an intrinsic identification problem that arises in using measures of productivity in tasks to infer traits.

• Analysts have to make one normalization in order to identify the traits.

• Need only one such construct joined with patterned structures on how \( \theta \) enters other task to identify the vector \( \theta \) (e.g., one example is a recursive, triangular structure).
Examples of Nonidentification Problems

IQ and Achievement Test Scores Reflect Incentives and Efforts, and Capture Both Cognitive and Personality Traits
Table 11: Incentives and Performance on Intelligence Tests

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample and Study Design</th>
<th>Experimental Group</th>
<th>Effect size of incentive (in standard deviations)</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edlund [1972]</td>
<td>Between subjects study. 11 matched pairs of low SES children; children were about one standard deviation below average in IQ at baseline</td>
<td>M&amp;M candies given for each right answer</td>
<td>Experimental group scored 12 points higher than control group during a second testing on an alternative form of the Stanford Binet (about 0.8 standard deviations)</td>
<td>“…a carefully chosen consequence, candy, given contingent on each occurrence of correct responses to an IQ test, can result in a significantly higher IQ score.” (p. 319)</td>
</tr>
<tr>
<td>Breuning and Zella [1978]</td>
<td>Within and between subjects study of 485 special education high school students all took IQ tests, then were randomly assigned to control or incentive groups to retake tests. Subjects were below-average in IQ.</td>
<td>Incentives such as record albums, radios (&lt;$25) given for improvement in test performance</td>
<td>Scores increased by about 17 points. Results were consistent across the Otis-Lennon, WISC-R, and Lorge-Thorndike tests.</td>
<td>“In summary, the promise of individualized incentives contingent on an increase in IQ test performance (as compared with pretest performance) resulted in an approximate 17-point increase in IQ test scores. These increases were equally spread across subtests… The incentive condition effects were much less pronounced for students having pretest IQs between 98 and 120 and did not occur for students having pretest IQs between 121 and 140.” (p. 225)</td>
</tr>
</tbody>
</table>

- Many other studies (see Almlund et al., 2011).
Factor Analysis: A Key Tool in Defining and Measuring Personality
Big 5: Constructed from Factor Analysis of a Variety of Self-Reported Measures
• $T_{n,l}$: trait $l$ for person $n$.
• Use multiple measures on the same traits to control for measurement error.

**Dedicated Factor Case**

• $P_{n,l}^q$: $q$th measurement on trait $l$ for person $n$.
• The $q$th measurement of factor $l$ for person $n$ is

$$P_{n,l}^q = \mu_{l}^q + \lambda_{l}^q T_{n,l} + \epsilon_{n,l}^q,$$

$q = 1, \ldots, Q_l, n = 1, \ldots, N, l = 1, \ldots, L$
• More general case:

\[ P_{n,l}^q = \mu_l^q + (\lambda^q)' T_n + \epsilon_{n,l}^q, \quad q = 1, \ldots, Q_l. \]  

• \( \lambda^q \) is a vector with possibly as many as \( L \) nonzero components.

• The \( \epsilon_{n,l}^q \) are assumed to be independent of \( T_n \) and mutually independent within and across constructs (\( l \) and \( l' \) are two constructs).

• Cunha, Heckman and Schennach [2010] develop nonlinear factor models (nonlinear and nonparameteric).
Conventional psychometric validity of a collection of items or test scores for different constructs has three aspects.

**Discriminant Validity**

(a) Factor $T_l$ for construct $l$ is statistically independent of factor $T_{l'}$ for construct $l' \neq l$.

**Convergent Validity**

(b) A factor $T_l$ is assumed to account for the intercorrelations among the items or tests within a construct $l$.

(c) Item-specific and random error variance are low (intercorrelations among items are high within a cluster).
Predictive Validity

- An alternative criterion for validating measurement systems is based on the predictive power of the tests for real world outcomes, that is, on behaviors measured outside of the exam room or observer system.
Problems with Predictive Validity

1. All measurements of factor $T_{n,l}$ can claim incremental predictive validity as long as each measurement is subject to error ($\epsilon_{n,l}^q \neq 0$).

2. Reverse causality.

3. Especially problematic when interpreting contemporary correlations between personality measurements and outcomes.
• The problem of reverse causality is sometimes addressed by using early measures of traits determined well before the outcomes are measured to predict later outcomes.

• This approach is problematic if the traits the analyst seeks to identify evolve over time and the contemporary values of traits drive behavior.

• Trades a reverse causality problem with a version of an errors in variables problem.

• Early measures of the traits may be poor proxies for the traits that drive current measured behavior.
Link to
Factor Models: A Brief Digression
Dedicated Factor Models Not Required

Example.

\[ R_1 = \alpha_{11}\theta_1 + (0)\theta_2 + \varepsilon_1 \]  (only one dedicated measurement on \( \theta_1 \))

\[ R_2 = \alpha_{21}\theta_1 + \alpha_{22}\theta_2 + \varepsilon_2 \]

\[ \vdots \]

\[ R_T = \alpha_{T1}\theta_1 + \alpha_{T2}\theta_2 + \varepsilon_T \]

- The \( \theta_1 \) and \( \theta_2 \) are freely correlated
Theorem 1

\((Williams, 2011)\)

If \((\theta_1, \theta_2) \perp (\varepsilon_1, \ldots, \varepsilon_T)\)

\[\varepsilon_i \perp \varepsilon_j, \quad \forall \ i \neq j\]

\[\theta_1 \not\perp \theta_2\]

Model identified if we normalize (e.g.) \(\alpha_{11} = 1; \alpha_{22} = 1\) and set \(\alpha_{12} = 0\).
The Quantitative Importance of Measurement Error

- The share of error variance for proxies of cognition, personality and investment ranges from 1%–90%.
- Not accounting for measurement error produces downward-biased estimates of self-productivity effects and perverse estimates of investment effects.
Table 12: Share of Residual Variance in Measurements of Cognitive Skills Due to the Variance of Cognitive Factor (Signal) and Due to the Variance of Measurement Error (Noise)

![Graph showing the share of residual variance in cognitive skills measurements.](image)

Source: Cunha, Heckman and Schennach [2010].
Table 13: Share of Residual Variance in Measurements of Socioemotional Skills Due to the Variance of Socioemotional Factor (Signal) and Due to the Variance of Measurement Error (Noise)

Source: Cunha, Heckman and Schennach [2010].
Table 14: Share of Residual Variance in Measurements of Investments Due to the Variance of Investment Factor (Signal) and Due to the Variance of Measurement Error (Noise)

Source: Cunha, Heckman and Schennach [2010].
Faking

- “Faking” may corrupt measurements designed to proxy latent factors.
- There are at least two types of false responses:
  (a) those arising from impression management and
  (b) those arising from self-deception (Paulhus [1984]).
• Reference bias
Reference Bias

Figure 19: National Rank in Big Five Conscientiousness and Average Annual Hours Worked
Stability of Traits
Changing Preference Parameters and Psychological Traits?

- If they change, to what extent do environments and investments influence the developmental trajectories of personality traits?
• The malleability of personality can be defined and measured in several ways: Mean-level change refers to change over time in absolute levels of a trait and is measured by changes in scores over time.

• Rank-order change, in contrast, refers to changes in the ordinal ranking of a trait in a population and is measured by test-retest rank correlations.

• Cognitive abilities exhibit dramatic mean-level change from early childhood through adolescence, but, over the same period, strong rank-order stability.
A second useful dichotomy contrasts normative change, defined as changes that are typical of the average individual in a given population, and caused either by biological programming (ontogenic) or by predictable changes in social roles (sociogenic), and non-normative change, encompassing both intentional change, caused by deliberate, self-directed efforts, deliberately chosen changes in social roles and atypical life events (trauma, for example).
Mean Level Changes

- People typically become more socially dominant
Social Dominance

![Graph showing Social Dominance over age]

- Cumulative d Value vs Age
- Data points indicating an increasing trend in Social Dominance with age.
Figure 4a

Cumulative mean-level changes in personality across the life course

Note: Figure taken from Roberts, Walton and Viechtbauer (2006). Reprinted with permission of the authors. Social vitality and social dominance are aspects of Big Five extraversion. Cumulative d values represent total lifetime change in standard deviations.
Conscientiousness

Figure 4a
Cumulative mean-level changes in personality across the life course

Note: Figure taken from Roberts, Walton and Viechtbauer (2006). Reprinted with permission of the authors. Social vitality and social dominance are aspects of Big Five extraversion. Cumulative d values represent total lifetime change in standard deviations.
Openness to Experience

Figure 4a
Cumulative mean-level changes in personality across the life course

Note: Figure taken from Roberts, Walton and Viechtbauer (2006). Reprinted with permission of the authors. Social vitality and social dominance are aspects of Big Five extraversion. Cumulative d values represent total lifetime change in standard deviations.

Borghans, Duckworth, Heckman, and ter Weel

Social Personality Psychology and Economics
Figure 4a Cumulative mean-level changes in personality across the life course

Note: Figure taken from Roberts, Walton and Viechtbauer (2006). Reprinted with permission of the authors. Social vitality and social dominance are aspects of Big Five extraversion. Cumulative d values represent total lifetime change in standard deviations.
Figure 4a
Cumulative mean-level changes in personality across the life course

Note: Figure taken from Roberts, Walton and Viechtbauer (2006). Reprinted with permission of the authors. Social vitality and social dominance are aspects of Big Five extraversion. Cumulative d values represent total lifetime change in standard deviations.
Figure 20: Mean-Level Changes in Cognitive Skills Using a Longitudinal Analysis

![Graph showing mean-level changes in cognitive skills over age with different markers and trendlines for various skills such as Inductive Reasoning, Spatial Orientation, Perceptual Speed, Numeric Ability, Verbal Ability, and Verbal Memory.]
Figure 20: Mean-Level Changes in Cognitive Skills Using a Cross-Sectional Analysis
Figure 4c

Fluid intelligence decreases and crystallized intelligence increases across the lifespan

Note: Figure from Horn (1970). Used with permission of Elsevier.
Rank-Order Change in Cognitive and Personality Skills

- Figure 21a shows graphs of rank order stability of personality by age.
- Figure 21b shows rank order stability of IQ over broad age ranges.
Figure 21: Rank Order Stability: Personality by Age

$r = 0.7$
Figure 21: Rank Order Stability: IQ over Broad Age Ranges

![Graph showing rank order stability over broad age ranges with correlations for nonverbal IQ, verbal IQ, total IQ, and Stanford-Binet IQ](image-url)
**Sources of Information on Personality**

- Self reports
- Teacher (or third party) reports
- Behaviors
- Computer games
- Choice experiments as elicited in game theory and experiments
- Use school records / teacher reports (available from public school records)
- Grades much more predictive of performance in college than SAT
Measurement System for Skills $\theta$

Using Behaviors to Proxy Skills

Heckman, Humphries, and Veramendi (forthcoming in *JPE*, 2016)

Also Kautz and Zanoni (2015)
• Let $\theta^C$ and $\theta^{SE}$ denote the levels of cognitive and socio-emotional endowments.
• Suppose $\theta = (\theta^C, \theta^{SE})$.
• $\theta^C$ and $\theta^{SE}$ are allowed to be correlated.
• Let $t^C_{m,s}$ be the $m^{th}$ cognitive test score and $t^{C,SE}_{m,s}$ the $m^{th}$ measure influenced by both cognitive and socio-emotional endowments, all measured at schooling level $s$.
• Parallel to the treatment of the index and outcome equations, we assume linear measurement systems for these variables:

\[
\begin{align*}
    t^C_{m,s} &= X^C_{m,s} \beta^C_{m,s} + \theta^C \alpha^C_{m,s} + e^C_{m,s}, \\
    t^{C,SE}_{m,s} &= X^{C,SE}_{m,s} \beta^{C,SE}_{m,s} + \theta^C \tilde{\alpha}^C_{m,s} + \theta^{SE} \tilde{\alpha}^{SE}_{m,s} + e^{C,SE}_{m,s}.
\end{align*}
\]
The structure assumed in Equations (9) and (10) is identified even when allowing for correlated factors, if there is one measure that is a determinant of cognitive endowments \( t_{m,s}^C \) and at least four measures that load on both cognitive ability and socio-emotional ability, and conventional normalizations are assumed.
• Heckman et al. (2016) normalize the cognitive loading to one for the arithmetic reasoning ASVAB measure.
• Normalize the socio-emotional loading to one for the language arts grade measure.
• One ability measure assumed dedicated.
• If any particular ASVAB score is excluded, it does not substantively change the analysis.
• Course grades are assumed to load on both the cognitive and socio-emotional factors Borghans et al. (2011, 2016).
As discussed above, the identification strategy used in the paper requires one measure that loads exclusively on cognitive ability.

- We assume ASVAB tests only measure cognition.
- Subject-specific 9th grade GPA, educational choices, and early risky behavior are assumed to depend on both factors.
- We include violent behavior, smoking regularly by age 15, drinking regularly by age 15, ever smoking marijuana by age 15, and sexual intercourse by age 15 as early “outcomes” in our model.
- These do not inform the cognitive or socio-emotional factor but provide a robustness check of our interpretation of our factors and aiding in interpretation.
Using Performance on Tasks to Measure Traits

Link to HCEO Conference (October, 2015)
https://hceconomics.uchicago.edu/events/conference-measuring-and-assessing-skills-0

Link to LaMar Slides
Appendix
### Table 15: Log Wage Regressions on Personality, IQ, and Achievement Tests in the BCS

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-esteem</strong></td>
<td>0.059***</td>
<td>0.058***</td>
<td>0.066***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Locus of control</strong></td>
<td>0.160***</td>
<td>0.146***</td>
<td>0.082***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.026)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Disorganized</strong></td>
<td>-0.001</td>
<td>0.018</td>
<td>0.045</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anti-social</strong></td>
<td>0.034</td>
<td>0.031</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neurotic</strong></td>
<td>-0.102***</td>
<td>-0.097***</td>
<td>-0.094***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Introvert</strong></td>
<td>-0.013</td>
<td>-0.020</td>
<td>-0.013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.026)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BAS Matrices (IQ)</strong></td>
<td>0.131***</td>
<td>0.074***</td>
<td>-0.064**</td>
<td>0.266***</td>
<td>0.254***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.027)</td>
<td>(0.021)</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td><strong>BAS Achievement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.207***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
</tr>
<tr>
<td><strong>PCLT (Achievement)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FMT (Achievement)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ERT (Achievement)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.024</td>
<td>0.003</td>
<td>-0.004</td>
<td>-0.016</td>
<td>-0.018</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.021)</td>
<td>(0.022)</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>2,589</td>
<td>2,589</td>
<td>2,589</td>
<td>2,362</td>
<td>2,362</td>
<td>2,153</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.015</td>
<td>0.051</td>
<td>0.055</td>
<td>0.066</td>
<td>0.090</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Heckman Personality Psychology and Economics
Table 15: Log Wage Regressions on Personality, IQ, and Achievement Tests in the BCS (cont.)

<table>
<thead>
<tr>
<th></th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
<th>(11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-esteem</td>
<td>0.043*</td>
<td>0.046</td>
<td>0.059**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.035)</td>
<td>(0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus of control</td>
<td>0.106***</td>
<td>0.071*</td>
<td>0.157***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.039)</td>
<td>(0.036)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disorganized</td>
<td>0.026</td>
<td>0.097**</td>
<td>0.049</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.048)</td>
<td>(0.042)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-social</td>
<td>0.015</td>
<td>0.019</td>
<td>0.020</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.042)</td>
<td>(0.036)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurotic</td>
<td>-0.080***</td>
<td>-0.041</td>
<td>-0.071**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.039)</td>
<td>(0.033)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introvert</td>
<td>-0.038</td>
<td>-0.079**</td>
<td>-0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.038)</td>
<td>(0.034)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAS Matrices (IQ)</td>
<td>0.023</td>
<td>-0.012</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.040)</td>
<td>(0.033)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAS Achievement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCLT (Achievement)</td>
<td>0.149***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMT (Achievement)</td>
<td></td>
<td>0.387***</td>
<td>0.373***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.038)</td>
<td>(0.053)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERT (Achievement)</td>
<td></td>
<td></td>
<td></td>
<td>0.239***</td>
<td>0.133***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.027</td>
<td>0.006</td>
<td>0.010</td>
<td>0.009</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.033)</td>
<td>(0.034)</td>
<td>(0.030)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,153</td>
<td>1,071</td>
<td>1,071</td>
<td>1,339</td>
<td>1,339</td>
</tr>
<tr>
<td>R²</td>
<td>0.064</td>
<td>0.089</td>
<td>0.105</td>
<td>0.043</td>
<td>0.077</td>
</tr>
</tbody>
</table>
Table 16: Log Wages Regressed on IQ, Rosenberg, Rotter, and AFQT in the NLSY79

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>0.173***</td>
<td>0.143***</td>
<td>0.019</td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.042)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotter</td>
<td>0.165</td>
<td>0.088</td>
<td>−0.007</td>
<td>−0.007</td>
<td>(0.148)</td>
<td>(0.149)</td>
<td>(0.150)</td>
</tr>
<tr>
<td></td>
<td>(0.148)</td>
<td>(0.149)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenberg</td>
<td>0.473***</td>
<td>0.369**</td>
<td>0.275*</td>
<td>0.274*</td>
<td>(0.147)</td>
<td>(0.149)</td>
<td>(0.150)</td>
</tr>
<tr>
<td></td>
<td>(0.148)</td>
<td>(0.149)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFQT</td>
<td>0.251***</td>
<td>0.237***</td>
<td>0.225***</td>
<td>0.213***</td>
<td>(0.042)</td>
<td>(0.059)</td>
<td>(0.061)</td>
</tr>
<tr>
<td></td>
<td>(0.148)</td>
<td>(0.149)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>10.215***</td>
<td>9.895***</td>
<td>9.985***</td>
<td>10.107***</td>
<td>(0.041)</td>
<td>(0.100)</td>
<td>(0.107)</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.102)</td>
<td>(0.107)</td>
<td>(0.107)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>554</td>
<td>554</td>
<td>554</td>
<td>554</td>
<td>554</td>
<td>554</td>
<td>554</td>
</tr>
<tr>
<td>R²</td>
<td>0.033</td>
<td>0.025</td>
<td>0.046</td>
<td>0.060</td>
<td>0.061</td>
<td>0.066</td>
<td>0.066</td>
</tr>
</tbody>
</table>
Table 17: Log Wage Regressed on Cognitive Ability and Personality in MIDUS

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Abil.</td>
<td>0.131***</td>
<td>0.102***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.023)</td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>0.113***</td>
<td>0.109***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>0.002</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-0.074**</td>
<td>-0.065*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.035)</td>
<td></td>
</tr>
<tr>
<td>Agreeableness</td>
<td>-0.172***</td>
<td>-0.164***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>0.114***</td>
<td>0.100***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1651</td>
<td>1651</td>
<td>1651</td>
</tr>
<tr>
<td>R²</td>
<td>0.018</td>
<td>0.053</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Heckman Personality Psychology and Economics
**Figure 22:** Probability of Being a High School Graduate at Age 30 and Not Going on to Further Education, Males

Notes: The data are simulated from the estimates of the model and the NLSY79 sample. Higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws). Solid lines depict probability, and dashed lines, 2.5%-97.5% confidence intervals. The upper curve is the joint density. The two marginal curves (ii) and (iii) are evaluated at the mean of the trait not being varied.  
*Source:* Heckman, Stixrud and Urzua [2006, Figure 19].
Figure 22: Probability of Being a High School Graduate at Age 30 and Not Going on to Further Education, Males

Notes: The data are simulated from the estimates of the model and the NLSY79 sample. Higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws). Solid lines depict probability, and dashed lines, 2.5%-97.5% confidence intervals. The upper curve is the joint density. The two marginal curves (ii) and (iii) are evaluated at the mean of the trait not being varied.
Source: Heckman, Stixrud and Urzua [2006, Figure 19].
Figure 23: Probability of Being a 4-year-college Graduate or Higher at Age 30, Males

Notes: The data are simulated from the estimates of the model and the NLSY79 sample. Higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws). Solid lines depict probability, and dashed lines, 2.5%-97.5% confidence intervals. The upper curve is the joint density. The two marginal curves (ii) and (iii) are evaluated at the mean of the trait not being varied. 
Source: Heckman, Stixrud and Urzua [2006, Figure 21].
**Figure 23:** Probability of Being a 4-year-college Graduate or Higher at Age 30, Males

**Notes:** The data are simulated from the estimates of the model and the NLSY79 sample. Higher deciles are associated with higher values of the variable. The confidence intervals are computed using bootstrapping (200 draws). Solid lines depict probability, and dashed lines, 2.5%-97.5% confidence intervals. The upper curve is the joint density. The two marginal curves (ii) and (iii) are evaluated at the mean of the trait not being varied.

**Source:** Heckman, Stixrud and Urzua [2006, Figure 21].
Return to main text
Appendix for Decker et al.
Digression on the Relationship Between Economic Preferences Paper
Table 18: Spearman correlation structure experimental data set

<table>
<thead>
<tr>
<th></th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Neuroticism</th>
<th>LoC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.0388</td>
<td>0.0162</td>
<td>−0.0114</td>
<td>0.1077**</td>
<td>−0.0684</td>
<td>0.1063*</td>
</tr>
<tr>
<td>Risk</td>
<td>0.0027</td>
<td>−0.0486</td>
<td>0.0786*</td>
<td>0.0206</td>
<td>−0.0995**</td>
<td>0.0485</td>
</tr>
<tr>
<td>Pos. Reciprocity</td>
<td>0.1606***</td>
<td>0.0078</td>
<td>0.0177</td>
<td>0.2020***</td>
<td>0.0152</td>
<td>0.0441</td>
</tr>
<tr>
<td>Neg. Reciprocity</td>
<td>−0.0967*</td>
<td>−0.0221</td>
<td>0.0462</td>
<td>−0.083*</td>
<td>−0.0165</td>
<td>−0.1376**</td>
</tr>
<tr>
<td>Trust</td>
<td>0.1354***</td>
<td>−0.1198***</td>
<td>0.002</td>
<td>0.1696***</td>
<td>−0.002</td>
<td>−0.0648</td>
</tr>
<tr>
<td>Altruism</td>
<td>0.0969*</td>
<td>−0.0804</td>
<td>0.0034</td>
<td>0.2000***</td>
<td>0.0879*</td>
<td>0.0418</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Correlations between economic preferences and the Big Five were calculated using 394–477 observations. Correlations between economic preferences and Locus of Control were calculated using 254–315 observations. All measures are standardized.
Table 19: Spearman correlation structure representative experimental data

<table>
<thead>
<tr>
<th></th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Neuroticism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>−0.0199</td>
<td>−0.0737</td>
<td>−0.0764*</td>
<td>−0.0829*</td>
<td>−0.0598</td>
</tr>
<tr>
<td>Risk</td>
<td>0.1315*</td>
<td>−0.0744</td>
<td>0.0661</td>
<td>−0.0854*</td>
<td>−0.0261</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. All measures are standardized.
### Table 20: Spearman Correlation Structure SOEP

<table>
<thead>
<tr>
<th></th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Neuroticism</th>
<th>LoC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>0.0233</td>
<td>0.1192</td>
<td>−0.0342</td>
<td>0.3099</td>
<td>−0.0643</td>
<td>0.0709</td>
</tr>
<tr>
<td>Risk</td>
<td>0.2632</td>
<td>−0.0500</td>
<td>0.2452</td>
<td>−0.1496</td>
<td>−0.1049</td>
<td>0.1426</td>
</tr>
<tr>
<td>Pos. Reciprocity</td>
<td>0.1835</td>
<td>0.2622</td>
<td>0.1547</td>
<td>0.1947</td>
<td>0.0808</td>
<td>0.1041</td>
</tr>
<tr>
<td>Neg. Reciprocity</td>
<td>−0.0616</td>
<td>−0.1767</td>
<td>−0.0426</td>
<td>−0.3853</td>
<td>0.0572</td>
<td>−0.2257</td>
</tr>
<tr>
<td>Trust</td>
<td>0.1224</td>
<td>−0.0693</td>
<td>0.0523</td>
<td>0.0788</td>
<td>−0.1889</td>
<td>0.2012</td>
</tr>
<tr>
<td>Altruism</td>
<td>0.1693</td>
<td>0.1501</td>
<td>0.1602</td>
<td>0.2416</td>
<td>0.0860</td>
<td>0.0843</td>
</tr>
</tbody>
</table>

All correlations are significant at the 1% level and are calculated using 14,243 observations. All measures are standardized.
**Figure 24:** Kernel-weighted local linear polynomial regressions using experimental data

### Openness

- **Time**
- **Risk**
- **Pos. Recip.**
- **Neg. Recip.**
- **Trust**
- **Altruism**

### Conscientiousness

- **Time**
- **Risk**
- **Pos. Recip.**
- **Neg. Recip.**
- **Trust**
- **Altruism**

---

Heckman

Personality Psychology and Economics
Figure 24: Kernel-weighted local linear polynomial regressions using experimental data Cont’d
Figure 24: Kernel-weighted local linear polynomial regressions using experimental data Cont’d
Figure 25: Kernel-weighted local linear polynomial regressions using SOEP data
Figure 25: Kernel-weighted local linear polynomial regressions using SOEP data Cont’d
Figure 25: Kernel-weighted local linear polynomial regressions using SOEP data Cont’d
### Table 21: Outcome Regressions: Representative Experimental Data

<table>
<thead>
<tr>
<th>Life Outcomes</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>0.043***</td>
<td>0.123***</td>
<td>0.989***</td>
<td>-0.018***</td>
<td>0.667***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.017)</td>
<td>(0.162)</td>
<td>(0.004)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Conscientiousn.</td>
<td>0.038***</td>
<td>0.106***</td>
<td>0.565***</td>
<td>-0.014***</td>
<td>-0.182***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.017)</td>
<td>(0.161)</td>
<td>(0.004)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Extraversion</td>
<td>0.026***</td>
<td>0.134***</td>
<td>-1.201***</td>
<td>0.006*</td>
<td>-0.309***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.017)</td>
<td>(0.154)</td>
<td>(0.004)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>0.033***</td>
<td>0.139***</td>
<td>-1.288***</td>
<td>0.023***</td>
<td>-0.146***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.018)</td>
<td>(0.165)</td>
<td>(0.004)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>-0.140***</td>
<td>-0.186***</td>
<td>-1.009***</td>
<td>0.018***</td>
<td>-0.272***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.016)</td>
<td>(0.158)</td>
<td>(0.004)</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. All measures are standardized.
Table 21: Outcome Regressions: Representative Experimental Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LoC</td>
<td>0.105***</td>
<td>0.307***</td>
<td>1.899***</td>
<td>-0.043***</td>
<td>0.421***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.145)</td>
<td>(0.003)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Patience</td>
<td>0.024***</td>
<td>0.129***</td>
<td>-0.343**</td>
<td>0.001</td>
<td>-0.151***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.136)</td>
<td>(0.003)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Risk</td>
<td>0.131***</td>
<td>0.076***</td>
<td>0.415**</td>
<td>0.003</td>
<td>0.210***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.017)</td>
<td>(0.166)</td>
<td>(0.004)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Pos. Recip.</td>
<td>-0.035***</td>
<td>0.006</td>
<td>0.388***</td>
<td>-0.002</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.140)</td>
<td>(0.003)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Neg. Recip.</td>
<td>0.064***</td>
<td>0.039**</td>
<td>-0.329**</td>
<td>0.006*</td>
<td>-0.137***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.147)</td>
<td>(0.003)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. All measures are standardized.
Table 21: Outcome Regressions: Representative Experimental Data
Cont’d

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subj. Health</td>
<td>0.122***</td>
<td>0.308***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life Satisf.</td>
<td></td>
<td></td>
<td>1.763***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross Wage</td>
<td></td>
<td></td>
<td></td>
<td>-0.035***</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.587***</td>
</tr>
<tr>
<td>Years of Educ.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.015)</td>
<td>(0.145)</td>
<td>(0.003)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Altruism</td>
<td>0.070***</td>
<td>0.072***</td>
<td>-0.780***</td>
<td>0.005</td>
<td>0.084***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.016)</td>
<td>(0.152)</td>
<td>(0.003)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.300***</td>
<td>6.852***</td>
<td>16.100***</td>
<td>0.099***</td>
<td>12.346***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.014)</td>
<td>(0.131)</td>
<td>(0.003)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Observations</td>
<td>14,218</td>
<td>14,214</td>
<td>7,199</td>
<td>9,095</td>
<td>13,768</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.108</td>
<td>0.159</td>
<td>0.0919</td>
<td>0.0547</td>
<td>0.174</td>
</tr>
</tbody>
</table>

*, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. All measures are standardized.
Pearson correlation coefficients between preference measures and life outcomes using SOEP data. Trust always shows the strongest association with life outcomes. More trust and a higher willingness to take risk are always related to better life outcomes, e.g. better health and greater life satisfaction, whereas negative reciprocity is associated with less life satisfaction and lower wages. The number of observations available varies for the different life outcomes: subjective health (14,218), life satisfaction (14,214), gross wage (7,199), unemployed (9,095), years of education (13,768). Gross wage measures the gross hourly wage.
Figure 27: Correlation Coefficients Between Personality Measures and Life Outcomes Using SOEP Data

Pearson correlation coefficients between personality measures and life outcomes using SOEP data. The locus of control and neuroticism show the strongest associations with life outcomes. A more internal locus of control is always related to better outcomes (e.g. better health or more life satisfaction), whereas a higher degree of neuroticism is associated with lower wages or a higher probability of being unemployed. The number of observations available varies for the different life outcomes: subjective health (14,218), life satisfaction (14,214), gross wage (7,199), unemployed (9,095), years of education (13,768). Gross wage measures the gross hourly wage.
Table 22: Linear representation of outcome regressions

<table>
<thead>
<tr>
<th></th>
<th>Subjective Health (OLS)</th>
<th>Subjective Health (o. probit)</th>
<th>Life Satisfaction (OLS)</th>
<th>Life Satisfaction (o. probit)</th>
<th>Gross Wage(OLS)</th>
<th>Gross Wage(o. probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Big5</td>
<td>LoC</td>
<td>Pref</td>
<td>Big5-Pref</td>
<td>Big5-Pref-LoC</td>
<td>Big5</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>0.0561</td>
<td>0.0383</td>
<td>0.0688</td>
<td>0.0975</td>
<td>0.1075</td>
<td>0.0220</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>170.04</td>
<td>567.35</td>
<td>176.01</td>
<td>140.59</td>
<td>143.72</td>
<td>834.99</td>
</tr>
<tr>
<td>AIC</td>
<td>37833</td>
<td>38094</td>
<td>37641</td>
<td>37201</td>
<td>37043</td>
<td>37139</td>
</tr>
<tr>
<td>BIC</td>
<td>37878</td>
<td>38109</td>
<td>37694</td>
<td>37292</td>
<td>37142</td>
<td>37207</td>
</tr>
<tr>
<td></td>
<td>0.0899</td>
<td>0.0782</td>
<td>0.0917</td>
<td>0.1342</td>
<td>0.1588</td>
<td>0.0261</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>281.88</td>
<td>1206.91</td>
<td>240.08</td>
<td>201.27</td>
<td>224.67</td>
<td>1406.38</td>
</tr>
<tr>
<td>AIC</td>
<td>55038</td>
<td>55216</td>
<td>55012</td>
<td>54335</td>
<td>53926</td>
<td>52448</td>
</tr>
<tr>
<td>BIC</td>
<td>55083</td>
<td>55231</td>
<td>55065</td>
<td>54426</td>
<td>54024</td>
<td>52561</td>
</tr>
<tr>
<td></td>
<td>0.0361</td>
<td>0.0388</td>
<td>0.0456</td>
<td>0.0704</td>
<td>0.0919</td>
<td>-</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>54.97</td>
<td>291.20</td>
<td>58.31</td>
<td>50.57</td>
<td>61.71</td>
<td>-</td>
</tr>
<tr>
<td>AIC</td>
<td>55088</td>
<td>55088</td>
<td>55042</td>
<td>54857</td>
<td>54690</td>
<td>-</td>
</tr>
<tr>
<td>BIC</td>
<td>55102</td>
<td>55102</td>
<td>55090</td>
<td>54940</td>
<td>54779</td>
<td>-</td>
</tr>
</tbody>
</table>

For the ordinary-least-squares (OLS) models we calculate $R^2$, whereas for the ordinal models we calculate pseudo $R^2$. The joint significance of all coefficients is tested using the F-test (OLS) and the LR-test (ordinal models). All F- and LR-tests are significant at the 1% level. With regard to the Akaike information criterion (AIC) and Bayesian information criterion (BIC), the smallest value for each outcome regression is underlined. Note that the full model (including the Big 5, locus of control and preferences) is always chosen by both information criteria. The number of observations available varies for the different life outcomes: subjective health (14,218), life satisfaction (14,214), gross wage (7,199), unemployed (9,095 obs.), and years of education (13,768). Gross wage measures the gross hourly wage.
Table 22: Linear representation of outcome regressions Cont’d

<table>
<thead>
<tr>
<th></th>
<th>Unemployed (OLS)</th>
<th></th>
<th>Unemployed (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Big5</td>
<td>LoC</td>
<td>Pref</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>0.0191</td>
<td>0.0331</td>
<td>0.0245</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>36.34</td>
<td>312.13</td>
<td>39.05</td>
</tr>
<tr>
<td>AIC</td>
<td>3067</td>
<td>2932</td>
<td>3017</td>
</tr>
<tr>
<td>BIC</td>
<td>3110</td>
<td>2946</td>
<td>3067</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Years of Education (OLS)</th>
<th></th>
<th>Years of Education (o. probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Big5</td>
<td>LoC</td>
<td>Pref</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>0.0914</td>
<td>0.0525</td>
<td>0.1061</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>277.93</td>
<td>763.89</td>
<td>273.29</td>
</tr>
<tr>
<td>AIC</td>
<td>65506</td>
<td>66078</td>
<td>65282</td>
</tr>
<tr>
<td>BIC</td>
<td>65551</td>
<td>66093</td>
<td>65335</td>
</tr>
</tbody>
</table>
### Table 23: Outcome Regressions: Flexible Specification

<table>
<thead>
<tr>
<th>Subjective Health (OLS)</th>
<th>Subjective Health (o. probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>adj. $R^2$/pseudo $R^2$</td>
</tr>
<tr>
<td>.0632 / .0388</td>
<td>.0251 / .0146</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>F-Test/LR-Test</td>
</tr>
<tr>
<td>48.99 / 288.17</td>
<td>952.98 / 555.19</td>
</tr>
<tr>
<td>AIC</td>
<td>AIC</td>
</tr>
<tr>
<td>37740</td>
<td>37051</td>
</tr>
<tr>
<td>BIC</td>
<td>BIC</td>
</tr>
<tr>
<td>37899</td>
<td>37232</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Life Satisfaction (OLS)</th>
<th>Life Satisfaction (o. probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>adj. $R^2$/pseudo $R^2$</td>
</tr>
<tr>
<td>.0948 / .0783</td>
<td>.0278 / .0219</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>F-Test/LR-Test</td>
</tr>
<tr>
<td>75.47 / 605.45</td>
<td>1493.78 / 1178.45</td>
</tr>
<tr>
<td>AIC</td>
<td>AIC</td>
</tr>
<tr>
<td>54976</td>
<td>52391</td>
</tr>
<tr>
<td>BIC</td>
<td>BIC</td>
</tr>
<tr>
<td>55135</td>
<td>52617</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gross Wage (OLS)</th>
<th>Gross Wage (OLS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>adj. $R^2$/pseudo $R^2$</td>
</tr>
<tr>
<td>.0382 / .0387</td>
<td>- / -</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>F-Test/LR-Test</td>
</tr>
<tr>
<td>15.30 / 145.74</td>
<td>- / -</td>
</tr>
<tr>
<td>AIC</td>
<td>AIC</td>
</tr>
<tr>
<td>55111</td>
<td>52617</td>
</tr>
<tr>
<td>BIC</td>
<td>BIC</td>
</tr>
<tr>
<td>55256</td>
<td>55298</td>
</tr>
</tbody>
</table>
### Table 23: Outcome Regressions: Flexible Specification Cont’d

<table>
<thead>
<tr>
<th>Subjective Health (OLS)</th>
<th>Unemployed (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big5</td>
<td>LoC</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>.0212</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>10.87</td>
</tr>
<tr>
<td>AIC</td>
<td>3062</td>
</tr>
<tr>
<td>BIC</td>
<td>3211</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Life Satisfaction (OLS)</th>
<th>Unemployed (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big5</td>
<td>LoC</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>.0948</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>75.47</td>
</tr>
<tr>
<td>AIC</td>
<td>54976</td>
</tr>
<tr>
<td>BIC</td>
<td>55135</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gross Wage (OLS)</th>
<th>Unemployed (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big5</td>
<td>LoC</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>-</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>15.30</td>
</tr>
<tr>
<td>AIC</td>
<td>-</td>
</tr>
<tr>
<td>BIC</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unemployed (OLS)</th>
<th>Unemployed (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big5</td>
<td>LoC</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>.0212</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>10.87</td>
</tr>
<tr>
<td>AIC</td>
<td>3062</td>
</tr>
<tr>
<td>BIC</td>
<td>3211</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unemployed (OLS)</th>
<th>Unemployed (probit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big5</td>
<td>LoC</td>
</tr>
<tr>
<td>adj. $R^2$/pseudo $R^2$</td>
<td>.0212</td>
</tr>
<tr>
<td>F-Test/LR-Test</td>
<td>10.87</td>
</tr>
<tr>
<td>AIC</td>
<td>3062</td>
</tr>
<tr>
<td>BIC</td>
<td>3211</td>
</tr>
</tbody>
</table>

The outcome variables are regressed on the indicated personality and preference measures. The difference with regard to the linear specification is that the model includes squares of all variables as well as all cross-products. Cross-products are also calculated between concepts in case more than one concept is included, e.g., in the Big 5-preferences case, we also include the cross-term neuroticism*risk. For the ordinary-least-squares (OLS) models we calculate $R^2$, whereas for the ordinal models we calculate pseudo-$R^2$. The joint significance of all coefficients is tested using the F-test (OLS models) and the LR-test (ordinal models). All F- and LR-tests are significant at the 1% level. With regard to the Akaike information criterion (AIC) and Bayesian information criterion (BIC), the smallest value for each outcome regression is underlined. Note that the full model (including the Big 5, locus of control and preferences) is chosen by both information criteria in nearly all cases; only for gross wage and unemployment does the BIC indicate that the model with only LoC and LoC$^2$ included should be used. The number of observations available varies for the different life outcomes: subjective health (14,218), life satisfaction (14,214), gross wage (7,199), unemployed (9,095), and years of education (13,768). Gross wage measures the gross hourly wage.
Return to main text
Appendix for Psychological Variables as Constraints
• A constraint-driven model need not produce a unique choice outcome for all persons with the same constraints.
• Thurstone (1927), Block and Marschak (1960), Bock and Jones (1968), and McFadden (1974, 1981), write the utility of agent $i$ for choice $l$ as $U_{i,l}$.

• $U_{i,l}$ is the motivation for choice (goal) $l$ by agent $i$.

• Choice sets, $B_i$, differ among persons depending on their capacities.

• Agent $i$ chooses $\hat{l}_i$ as the maximal element in the choice set $B_i$:

$$\hat{l}_i = \arg \max_{l \in B_i} \{U_{i,l}\}$$
A familiar model writes $U_{i,l} = V_{i,l} + \varepsilon_{i,l}$, where $V_{i,l}$ is agent $i$ valuation for $l$ and $\varepsilon_{i,l}$ is a random “taste” shock.

When $V_{i,l} = V_l$, and $\varepsilon_{i,l}$ is iid extreme value type 1, the probability that $l$ is selected from choice set $B_i$ is

$$\Pr(l \mid B_i) = \frac{\exp(V_l)}{\sum_{j \in B_i} \exp(V_j)} \quad \text{for } l \in B_i$$

$$= 0, \text{ for } l \notin B_i. \quad (11)$$

If agents have zero mean scale preference among the choices ($V_l = 0$) so that all choices (goals) have the same mean utility, we obtain a version of Becker’s (1962) model of irrational behavior.
Depending on how the constraints are determined, one can capture a variety of aspects of choice behaviour.

- A shy person may limit her options in a way an extrovert does not.
- An intelligent person may have a much richer choice set not only because of greater earnings capacity but also because of much greater imagination.
- Much like greater pixel resolution in imaging machines, those with higher IQ may resolve reality in a more fine-grained and less biased way.
- We capture the effect of these traits on the choice sets, which may also depend on material endowments.
Return to main text
Appendix for A Third Model
• How should one incorporate psychological traits into conventional economic models?
• One could think of them as public goods.
• This is the approach implicitly adopted by most personality psychologists.
• One could also think of psychological traits as excludable private goods.
• More of a trait used in one activity means less of the trait available for use in other activities.
• In addition, one might augment, complement or override the supply of a trait to any activity by supplying more time, or energy, to the activity in which the trait is used.

• On the other hand, “energy,” \( e \), which can be vector valued, may be used to moderate the manifestation of the trait (for example, energy may be spent controlling anger in a given activity).

• Individuals differ in their endowment vector of the trait \( \bar{\theta} \).

• Thus there may be a time constraint as in Becker (1965) or more generally there may be energy constraints (constraints on effort capacity).
A One-period Model

- Assume that there are $J$ activities with outputs $Z_j$, $j = 1, \ldots, J + 1$.
- We add one activity to account for market earnings. $Z_j$ is produced by combining tasks, $T_j$, defined in section II, with purchased market goods, $X_j$. 
• Augment the task functions defined by Equation (11) to include levels of energy, and time, in vector $e^j$

$$T_j = h_j (\theta^j, e^j) \text{ for } j = 1, \ldots, J + 1$$ (12)

$\theta^j$ is to be distinguished from $\theta_j$, the $j^{\text{th}}$ component of vector $\theta$.

• Parallel notation for $e^j$.

• For a fixed input of psychological traits, higher levels of $e^j$ may raise the output of the task.

• Thus if $e^j = 0$, the trait $\theta^j$ may be switched off. However, if some traits have negative productivity in some tasks more energy may be allocated to those tasks to offset the negative trait.
Output in activity $Z_j$ is

$$Z_j = \varphi_j (T_j, X_j) \text{ for } j = 1, \ldots, J + 1$$  \hfill (13)

- The outputs in activity $j$ depend on the task output $T_j$ and the goods input $X_j$.
- Agents have preferences over $Z_j$ and $e_j$.
- The effort expended in an activity may have psychic costs or benefits.
- There may be psychic costs in using $e_j$ to suppress the expression of a trait.
• Preferences may also depend on $\theta$ as well as other variables which we keep implicit.

• The utility function is

$$U = U (Z_1, \ldots, Z_j, e^1, \ldots, e^{J+1}, \theta) \quad (14)$$

• Income is return on asset flow $Y$ plus labor earnings which we denote $Z_{J+1} = \varphi_{J+1} (T_{J+1}, X_{J+1})$.

$$\sum_{j=1}^{J+1} P_j X_j = Y + Z_{J+1} \quad (15)$$

• $Z_{J+1}$ is a hedonic earnings function which prices out traits and energy in the market.
Two Ways to Introduce $\theta$

- It is possible to distinguish two different cases for $\theta$.
- For psychological traits, we can distinguish the case where $\theta$ is a public good, $\theta^j = \bar{\theta}$ for all $j = 1, \ldots, J+1$.
- When it is a private good, $\sum_{j=1}^{J+1} \theta^j = \bar{\theta}$
- People are not stuck with their personality in all activities.
For simplicity, we consider the pure private goods case and the pure public goods case. Assume that $e$ is private.

<table>
<thead>
<tr>
<th></th>
<th>Public</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e$ Private</td>
<td>case I</td>
<td>case II</td>
</tr>
</tbody>
</table>

In case I, the additional constraint operating on the consumer beyond the budget constraint (15) is

$$\theta^j = \bar{\theta}, \quad \sum_{j=1}^{J+1} e^j = \bar{e}, \text{ for all } j = 1, \ldots, J + 1. \quad (16)$$
• In case II, the operative constraints are

\[ \sum_{j=1}^{J+1} \theta^j = \bar{\theta}, \quad \sum_{j=1}^{J+1} e^j = \bar{e} \]

(17)
Case I: Traits as Public Goods

- In case I, different bundles of $\bar{\theta}$ across persons create comparative advantages for agents in different tasks and thus produce comparative advantages in different activities.
- Case I is a version of Michael’s (1973) model of environmental variables in a household production framework.
• For analytical simplicity, suppose that $Z_j$ and $T_j$, $j = 1, \ldots, J + 1$, display constant returns to scale in non-public inputs.
• In terms of the technologies (11), when \( \theta \) is a public good, we assume constant returns to scale in \( e^j \) but that \( \theta^j = \bar{\theta} \) is a fixed, environmental variable.

• Different levels of \( \bar{\theta} \) produce different productivities in different tasks.

• Feeding \( \bar{\theta} \) into the activity functions (13), which are also assumed to be constant returns to scale, we can analyze the agent’s problem of allocating effort among tasks and goods among activities using the analysis of Michael (1973).

• Financial and energy resources are not changed by \( \bar{\theta} \) except for its effect on \( Z_{J+1} \).

• Holding energy and money resources fixed, changes in \( \bar{\theta} \) produce reallocations across budget categories.
Several Cases

- Consider an increase in conscientiousness.
- This will likely increase earnings (via $Z_{J+1}$), and will enhance productivity in some tasks intensive in conscientiousness and activities based on those tasks more than other tasks and activities.
- The increased income will support more of all activities.
- The differential shift in productivity across tasks and activities will reduce the prices of activities that are more intensive in the use of conscientiousness.
- If the demands for those activities are price elastic compared to the demands for the less conscientiousness-intensive activities, the demand for the inputs used in those activities will increase.
- If the demands are relatively inelastic, the demands will decrease because of the greater productivity for the inputs.
- Standard Marshall’s 4 rules analysis.
• If a trait reduces productivity, the chain of logic just presented runs in reverse.
• With increases in (for example) neuroticism, shadow prices of activities intensive in that trait will increase.
• Labor earnings will tend to decrease.
• In the price-elastic case, consumers will tend to substitute away from activities intensive in the trait and the demand for inputs will decrease.
• In the inelastic case, input demands will increase as agents substitute goods and energy inputs into the activities that are inelastically demanded.
• The same level of the traits is found in all activities, but in general, energy or time will be allocated differentially among activities.

• A person who allocates more energy or time to a task will manifest more of the trait.

• If inputs are complementary, at the same scale of output more of the task will be demanded.

• Unless one controls for these inputs, one may fail to capture the uniformity of traits across tasks and activities.

• In all of these cases, purchase patterns of market goods will provide information on endowments and allocation of energy and traits.
Case II: **Traits as Private Goods**

- The case when traits are private goods produces the possibility of different levels of traits being used in different tasks and activities.

- Responses of activity levels to changes in rewards across activities will be more price-elastic when traits can be allocated across activities than when traits are fixed.

- Equiproportionate expansions in $(\bar{\theta}, \bar{e})$ differentially expand the consumption possibility set for activities intensive in $(\theta, e)$ and reduce their shadow prices, producing substitution effects in task production and activity consumption that promote consumption in activities intensive in the traits.
• The public goods case imposes more constraints on the system than the private goods case.

• Compared to the case of public goods for traits, agents will reduce their allocation of the trait from activities where their productivity is negative and will spend less effort \((e)\) in overriding the effects of negative traits in productivity.

• The trait will be shifted into less costly activities and less energy will be spent controlling it.
The Evidence

- The evidence would seem to favor case II, since different levels of traits are often found in different activities.
- However, since most of the estimates reviewed do not adjust for the inputs that affect the manifestation of the traits, one must be cautious in reaching this conclusion.
- Such adjustments are indicated by the theory but are not yet standard in economics or psychology.
• The roles of time and energy in amplifying or reducing the effects of the traits in activities needs to be systematically explored to make the theory empirically operational as are the effects of traits on the purchase of related goods (for example, shy people may seek to live in secluded areas, houses with high walls and seek jobs with little human contact).

• In the private goods specification of the model (case II), the motivation for the supply of traits to different activities depends on preferences (utility rewards $U$), on productivity in $Z_j$, and in productivity in the tasks $T_j$. In this framework, it is possible to formalize many of the currently disparate concepts of personality psychology.

• It would be very informative to estimate both versions of the model and to test between them.
Return to main text
Altruism and Social Preferences
• There is a large literature in economics on altruism and an emerging literature in economics on social preferences.
• Bergstrom (1997) and Laitner (1997) discuss models of interdependent family preferences.
• Andreoni (1995) shows that pure models of altruism are inconsistent with the evidence ("warm glow").
• Villanueva (2005) and Laferrre and Wolff (2006) summarize the mixed evidence on altruism in families.
A recent literature explores social preferences which are distinct from altruism per se.

Altruism is based on the assumption that the preferences of one agent depend on the consumption or utility of other agents.

Social preferences are preferences that depend on agent’s evaluations of a social condition (inequality, for example) or the intentions of other agents.

Fehr and Schmidt (1999) analyze inequality aversion (in which people dislike inequality rather than valuing the consumption or utility of agents per se).

Fehr and Gachter (2000), and Falk and Fischbacher (2006) present evidence on reciprocity and conditional cooperation, in which agents act in a pro-social or antisocial manner depending on the behavior of others with whom they interact.

Fehr and Schmidt (2006) summarize the theory and empirical support for social preferences.
Return to main text
Additional Validity Studies
### Table 24: Predictive Validities in Outcomes that Matter (Adjusted R-Squared)

<table>
<thead>
<tr>
<th></th>
<th>IQ Sample</th>
<th>AFQT Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
<td>Personality</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.01</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Adj, $R^2$ Cog, Personality 0.07 0.17

**Notes:** † Uses mean GED subtest scores  
‡ Uses a general GED factor
**Table 24: Predictive Validities in Outcomes that Matter (Adjusted R-Squared) (cont.)**

<table>
<thead>
<tr>
<th>Males</th>
<th>GPA Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GPA</td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.09</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.07</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.02</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.03</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.01</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.03</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.14</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Adj, R^2 Cog, Personality</strong></td>
<td><strong>0.11</strong></td>
</tr>
</tbody>
</table>

**Notes:** † Uses mean GED subtest scores
‡ Uses a general GED factor
Table 24: Predictive Validities in Outcomes that Matter (Adjusted R-Squared) (cont.)

<table>
<thead>
<tr>
<th>Females</th>
<th>IQ Sample</th>
<th>AFQT Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IQ</td>
<td>Personality</td>
</tr>
<tr>
<td>Earnings at Age 35</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>-0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.02</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Adj, $R^2$ Cog, Personality 0.10 0.15

Notes: † Uses mean GED subtest scores
‡ Uses a general GED factor
Table 24: Predictive Validities in Outcomes that Matter (Adjusted R-Squared) (cont.)

<table>
<thead>
<tr>
<th>Females</th>
<th>GPA Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Earnings at Age 35</td>
<td>0.05 0.04 0.07</td>
</tr>
<tr>
<td>Hourly Wage at Age 35</td>
<td>0.06 0.04 0.08</td>
</tr>
<tr>
<td>Hours Worked at Age 35</td>
<td>0.00 0.01 0.01</td>
</tr>
<tr>
<td>Jail by Age 35</td>
<td>0.01 0.01 0.02</td>
</tr>
<tr>
<td>Welfare at Age 35</td>
<td>0.05 0.05 0.07</td>
</tr>
<tr>
<td>Married at Age 35</td>
<td>0.03 0.03 0.05</td>
</tr>
<tr>
<td>BA Degree by Age 35</td>
<td>0.10 0.08 0.13</td>
</tr>
<tr>
<td>Depression in 1992</td>
<td>0.02 0.05 0.05</td>
</tr>
<tr>
<td>Adj, $R^2$ Cog, Personality</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Notes: † Uses mean GED subtest scores
‡ Uses a general GED factor
Return to main text
Factor Models: A Brief Digression
Suppose we have five measurements on $R_i$

- $i = 1, \ldots, 5$
- $\tilde{R}_i = \mu_i + R_i$
- $E(R_i) = 0$
- Then $E(\tilde{R}_i) = \mu_i$ and we identify means of measurements
Identifying Variances and Factor Loadings

\[ R_1 = \alpha_1 \theta + \varepsilon_1, \quad R_2 = \alpha_2 \theta + \varepsilon_2, \quad R_3 = \alpha_3 \theta + \varepsilon_3, \]
\[ R_4 = \alpha_4 \theta + \varepsilon_4, \quad R_5 = \alpha_5 \theta + \varepsilon_5, \]
\[ \varepsilon_i \perp \perp \varepsilon_j, \quad i \neq j, \quad \theta \perp \varepsilon_i, \quad i = 1, \ldots, 5 \]

\[ E(\theta) = 0; \quad E(\varepsilon_i) = 0; \quad i = 1, \ldots, 5 \]

\[ \text{Cov}(R_1, R_2) = \alpha_1 \alpha_2 \sigma_\theta^2 \]
\[ \text{Cov}(R_1, R_3) = \alpha_1 \alpha_3 \sigma_\theta^2 \]
\[ \text{Cov}(R_2, R_3) = \alpha_2 \alpha_3 \sigma_\theta^2 \]

- Normalize \( \alpha_1 = 1 \)

\[ \frac{\text{Cov}(R_2, R_3)}{\text{Cov}(R_1, R_2)} = \alpha_3 \]
\begin{itemize}

\item \[ \therefore \text{We know } \sigma^2_\theta \text{ from } \text{Cov} (R_1, R_2). \]
\item From \[ \text{Cov} (R_1, R_3) \] we know \[ \alpha_3, \alpha_4, \alpha_5. \]
\item Can get the variances of the \[ \varepsilon_i \] from variances of the \[ R_i \]
\[ \text{Var}(R_i) = \alpha_i^2 \sigma^2_\theta + \sigma^2_{\varepsilon_i}. \]
\item If \[ T = 2 \], all we can identify is \[ \alpha_1 \alpha_2 \sigma^2_\theta. \]
\item If \[ \alpha_1 = 1 \], and \[ \sigma^2_\theta = 1 \], we identify \[ \alpha_2. \]
\item Otherwise model is fundamentally underidentified.
\item One factor model requires three (or more) measurements, plus a normalization (to set scale)
\end{itemize}
2 Factors: (some examples)

\[
\theta_1 \perp \theta_2
\]

(For example.) This is not required in general (but it is for this example).

\[
\varepsilon_i \perp \varepsilon_j \quad \forall i \neq j
\]

\[
R_1 = \alpha_{11}\theta_1 + (0)\theta_2 + \varepsilon_1
\]
\[
R_2 = \alpha_{21}\theta_1 + (0)\theta_2 + \varepsilon_2
\]
\[
R_3 = \alpha_{31}\theta_1 + \alpha_{32}\theta_2 + \varepsilon_3
\]
\[
R_4 = \alpha_{41}\theta_1 + \alpha_{42}\theta_2 + \varepsilon_4
\]
\[
R_5 = \alpha_{51}\theta_1 + \alpha_{52}\theta_2 + \varepsilon_5
\]

Let \( \alpha_{11} = 1, \alpha_{32} = 1 \). (Set scale)
\[ \text{Cov} (R_1, R_2) = \alpha_{21} \sigma_{\theta_1}^2 \]
\[ \text{Cov} (R_1, R_3) = \alpha_{31} \sigma_{\theta_1}^2 \]
\[ \text{Cov} (R_2, R_3) = \alpha_{21} \alpha_{31} \sigma_{\theta_1}^2 \]

- Form ratio of \[ \frac{\text{Cov} (R_2, R_3)}{\text{Cov} (R_1, R_2)} = \alpha_{31} \]
- \[ \therefore \text{we identify } \alpha_{31}, \alpha_{21}, \sigma_{\theta_1}^2 \]

\[ \text{Cov} (R_1, R_4) = \alpha_{41} \sigma_{\theta_1}^2, \quad \therefore \text{since we know } \sigma_{\theta_1}^2 \therefore \text{we get } \alpha_{41}. \]

\[ \vdots \]
\[ \text{Cov} (R_1, R_k) = \alpha_{k1} \sigma_{\theta_1}^2 \]

- \[ \therefore \text{we identify } \alpha_{k1} \text{ for all } k \text{ and identify } \sigma_{\theta_1}^2. \]
Can Identify Other Parameters

\[
\text{Cov}(R_3, R_4) - \alpha_{31} \alpha_{41} \sigma_{\theta_1}^2 = \alpha_{42} \sigma_{\theta_2}^2 \\
\text{Cov}(R_3, R_5) - \alpha_{31} \alpha_{51} \sigma_{\theta_1}^2 = \alpha_{52} \sigma_{\theta_2}^2 \\
\text{Cov}(R_4, R_5) - \alpha_{41} \alpha_{51} \sigma_{\theta_1}^2 = \alpha_{52} \alpha_{42} \sigma_{\theta_2}^2,
\]

• By similar logic,

\[
\frac{\text{Cov}(R_4, R_5) - \alpha_{41} \alpha_{51} \sigma_{\theta_1}^2}{\text{Cov}(R_3, R_4) - \alpha_{31} \alpha_{41} \sigma_{\theta_1}^2} = \alpha_{52}
\]

• ∴ we also identify \( \sigma_{\theta_2}^2 \) for “2” loadings.
• If we have dedicated measurements on each factor do not need normalizations on the factors of $R$.
• Dedicated measurements set the scales and make factor models interpretable:

$$M_1 = \theta_1 + \varepsilon_{1M}$$
$$M_2 = \theta_2 + \varepsilon_{2M}$$

$$\text{Cov} (R_1, M) = \alpha_{11} \sigma_{\theta_1}^2$$
$$\text{Cov} (R_2, M) = \alpha_{21} \sigma_{\theta_1}^2$$
$$\text{Cov} (R_3, M) = \alpha_{31} \sigma_{\theta_1}^2$$

$$\text{Cov} (R_1, R_2) = \alpha_{11} \alpha_{21} \sigma_{\theta_1}^2$$
$$\text{Cov} (R_1, R_3) = \alpha_{11} \alpha_{31} \sigma_{\theta_1}^2$$
• Form the ratio

$$\frac{\text{Cov}(R_1, R_2)}{\text{Cov}(R_1, M)} = \alpha_{12}$$

∴ We identify $\sigma_{\theta_1}^2$, etc.

• ∴ We can identify $\alpha_{12}, \sigma_{\theta_1}^2$ and the other factors.
General Case

\[ R_{T \times 1} = \mu_{T \times 1} + \Lambda_{T \times K} \theta_{K \times 1} + \varepsilon_{T \times 1} \]

- \( \theta \) are factors, \( \varepsilon \) uniquenesses

\[ E(\varepsilon) = 0 \]

\[ \text{Var}(\varepsilon \varepsilon') = D = \begin{pmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma^2_{\varepsilon_T} \end{pmatrix} \]

\[ E(\theta) = 0 \]

\[ \text{Var}(R) = \Lambda \Sigma_\theta \Lambda' + D \]

\[ \Sigma_\theta = E(\theta \theta') \]
• The only source of information on $\Lambda$ and $\Sigma_\theta$ is from the covariances.
• Each variance is “contaminated” by a uniqueness.
• Associated with each variance of $R_i$ is a $\sigma^2_{\varepsilon_i}$.
• Each uniqueness variance contributes one new parameter.
• How many unique covariance terms do we have?
  \[ T(T - 1) \]
• \[ \frac{T(T - 1)}{2} \]
• We have $T$ uniquenesses; $TK$ elements of $\Lambda$.

\[
\frac{K(K-1)}{2}
\]

• $\frac{K(K-1)}{2}$ elements of $\Sigma_\theta$.

• $\frac{K(K-1)}{2} + TK$ parameters ($\Sigma_\theta, \Lambda$).

• Need this many covariances to identify model “Ledermann Bound”:

\[
\frac{T(T-1)}{2} \geq TK + \frac{K(K-1)}{2}
\]

• ($\# \ of \ equations \ \geq \ # \ of \ unknowns.$)
Lack of Identification Up to Rotation

- Observe that if we multiply $\Lambda$ by an orthogonal matrix $C$, $(CC' = I)$, we obtain

$$Var (R) = \Lambda C [C' \Sigma_{\theta} C] C' \Lambda' + D$$

- $C$ is a “rotation.”
- Cannot separate $\Lambda C$ from $\Lambda$.
- Model not identified against orthogonal transformations in the general case.
Some common assumptions:

(i) $\theta_i \perp \perp \theta_j$, $\forall i \neq j$

$$
\Sigma_{\theta} = \begin{pmatrix}
\sigma^2_{\theta_1} & 0 & \ldots & 0 \\
0 & \sigma^2_{\theta_2} & 0 & \vdots \\
\vdots & 0 & \ddots & \vdots \\
0 & \ldots & 0 & \sigma^2_{\theta_K}
\end{pmatrix}
$$
joined with

(ii)

\[ \Lambda = \begin{pmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
\alpha_{21} & 0 & 0 & 0 & \ldots & 0 \\
\alpha_{31} & 1 & 0 & 0 & \ldots & 0 \\
\alpha_{41} & \alpha_{42} & 0 & 0 & \ldots & 0 \\
\alpha_{51} & \alpha_{52} & 1 & 0 & \ldots & 0 \\
\alpha_{61} & \alpha_{62} & \alpha_{63} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & 1 & \vdots \\
\end{pmatrix} \]
Return to main text
Models for Understanding Student Thinking using Data from Complex Computerized Science Tasks
Michelle M. LaMar (2014)
Using Markov Decision Processes to Understand Student Thinking in Performance Tasks

Michelle M. LaMar
Educational Testing Service
mlamar@ets.org

October 1st, 2015
23. Within a substance, atoms that collide frequently and move independently of one another are most likely in a

A. liquid.
B. solid.
C. gas.
D. crystal.

Traditional Assessment Task

Standard Educational Measurement Paradigm
23. Within a substance, atoms that collide frequently and move independently of one another are most likely in a

A. liquid.
B. solid.
C. gas.
D. crystal.

Traditional Assessment Task

Not very similar

Real Science Task
23. Within a substance, atoms that collide frequently and move independently of one another are most likely in a

A. liquid.
B. solid.
C. gas.
D. crystal.

Simulated Task

Traditional Science Task
Assessing Science Skills

Give students some equipment and see what they do.
• What is their goal?
• How much do they care?
• Who is contributing? How much?
• Do they understand how the equipment works?
• Are they using good inquiry skills?
• Do they understand the science content?
Assessing Science Skills

Give students some equipment and see what they do.

- Goals
- Motivation
- Collaboration Skills
- Beliefs & Understanding of Task Setup
- Science Process Skills
- Science Content Knowledge
Assessing Science Skills

Give students a standard assessment item.

- Goals
- Motivation
- Collaboration
- Beliefs & Understanding
- Science Process
- Science Knowledge

23. Within a substance, atoms that collide frequently and move independently of one another are *most* likely in a

A  liquid.
B  solid.
C  gas.
D  crystal.
Assessing Science Skills

Give students some equipment and see what they do.

• Goals
• Motivation
• Collaboration
• Beliefs & Understanding
• Science Process
• Science Knowledge
Assessing Science Skills

Give students some equipment and see what they do.

- Goals
- Motivation
- Collaboration
- Beliefs & Understanding
- Science Process
- Science Knowledge

Model
All Factors of Interest
Latent-trait Models

Mislevy: The conditional probability model-fragments:

\[ p(X_{ijk} | \theta_i, \beta_j, \zeta_k) \]

- \( X_{ijk} \) is the “observable” variable from the action(s) of “Person i” in “Situation j” given other relevant contextual variables \( k \);
- \( \theta_i \) is the “proficiency” variable for “Person i” (might also subscript for time \( t \));
- \( \beta_j \) is the effect of “Situation j”; and
- \( \zeta_k \) is the effect of other relevant contextual variables \( k \).
Cognitive Process Models

Action choice based on human and environment:

\[ p(a_{ijk} \mid \theta_i, \beta_j, \zeta_k) \]

- \( a_{ijk} \) is the “observable” actions of “Person i” in “Situation j” given other relevant contextual variables \( k \);
- \( \theta_i \) is the “proficiency” variable for “Person i” (might also subscript for time \( t \));
- \( \beta_j \) is the effect of “Situation j”; and
- \( \zeta_k \) is the effect of other relevant contextual variables \( k \).
Outline

• Peg Solitaire Example
• Markov Decision Process Measurement Model
  • The MDP
  • The MDP for Measurement
• MDP-MM in Action
  • Peg Solitaire
  • Microbes
  • SimCityEDU Pollution Challenge
• Conclusions
Example: Peg Solitaire Game

- Goal: leave as few pegs on the board as possible
- Jump pegs to remove them
Example: Peg Solitaire Game

- Goal: leave as few pegs on the board as possible
- Jump pegs to remove them

Can we estimate student strategic ability from a single game play record?
Example: Peg Solitaire Game

[Diagram of a peg solitaire board with pegs and moves highlighted]
Example: Peg Solitaire Game

Process Data, Action Sequence:

(3,3) → (1,3)
(3,5) → (3,3)
(4,3) → (2,3)
(1,3) → (3,3)
(3,2) → (3,4)

Score
Example: Peg Solitaire Game

State Sequence
Example: Peg Solitaire Game

State Space

22 Reachable States
Example: Peg Solitaire Game

Each state presents a choice:

Available Actions $A_s$

- $(3,3) \rightarrow (1,3)$
- $(3,3) \rightarrow (5,3)$
- $(3,2) \rightarrow (3,4)$
- $(4,3) \rightarrow (2,3)$

Score
Reset

Want

\[
p(a|s, \theta_j) = f(\theta_j, \xi_s)
\]
Example: Peg Solitaire Game

Each state presents a choice:

Available Actions $A_s$

- $(3,3) \rightarrow (1,3)$
- $(3,3) \rightarrow (5,3)$
- $(3,2) \rightarrow (3,4)$
- $(4,3) \rightarrow (2,3)$

Score

Reset

Want $p(a|s, \theta_j) = f(\theta_j, \xi_s)$
Example: Peg Solitaire Game

Each state presents a choice:

Available Actions $A_s$

1. (3,3) → (1,3)
2. (3,3) → (5,3)
3. (3,2) → (3,4)
4. (4,3) → (2,3)

Score

Reset

Want $p(a|s, \theta_j) = f(\theta_j, \xi_s)$

Markov Decision Process
Markov Decision Process

• Model for sequential planning in the presence of uncertainty.

• Developed in the 1950s for process optimization in robotics (Bellman 1957).

• Recently used in cognitive science to model how we infer another person’s motivations and beliefs (Baker, Saxe, Tennenbaum, 2009)
Markov Decision Process

<table>
<thead>
<tr>
<th>State Space</th>
<th>$S = {s_1, s_2, \ldots, s_S}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action Set</td>
<td>$A = {a_1, a_2, \ldots, a_A}$</td>
</tr>
<tr>
<td>Transition Function</td>
<td>$T(s, a, s') = p(s'</td>
</tr>
<tr>
<td>Reward Structure</td>
<td>$R(s, a, s')$</td>
</tr>
</tbody>
</table>

Policy: $p(a|s, \xi)$
Markov Decision Process

\[ p(a|s, \xi) = f \text{(The value of action } a) \]
Markov Decision Process

\[ p(a|s, \xi) = f(\text{The value of action } a) \]

Expected Total Rewards
Markov Decision Process

\[ p(a|s, \xi) = f(\text{The value of action } a) \]
Markov Decision Process

\[ p(a|s, \xi) = f(\text{The value of action } a) \]
Markov Decision Process

The expected rewards for taking action \( a \) in state \( s \) is expressed by the Q-function (Bellman, 1957):

\[
Q(s, a) = \sum_{s' \in S} p(s'|s, a) \left( R(s, a, s') + \gamma \sum_{a' \in A} p(a'|s)Q(s', a') \right)
\]
Markov Decision Process

The expected rewards for taking action \( a \) in state \( s \) is expressed by the Q-function (Bellman, 1957):

\[
Q(s, a) = \sum_{s' \in S} p(s'|s, a) \left( R(s, a, s') + \gamma \sum_{a' \in A} p(a'|s') Q(s', a') \right)
\]

- **Value of choosing action** \( a \) in state \( s \)
- **Transition probability**
- **Immediate Reward**
- **Discounted Expected Future Reward**
Decision Process

In robotics, solve for the optimal policy:

$$\pi(s) \equiv \arg\max_{a \in A} (Q^*(s, a)), \quad p(a \in \pi(s)|s) = 1$$

In psychology, the Boltzmann policy is used

$$p(x_{sj} = a|s) \propto e^{\beta Q(s, a)}$$

$$\beta \in [0, \infty)$$

Consider $\beta_j$ as a person-specific “capability”

$$p(x_{sj} = a|\beta_j, s) \propto e^{\beta_j Q(s, a|\beta_j)}$$
MDP as a Measurement Model

Full MDP Measurement model:

\[ p(x_{sj} = a|s, \beta_j) = \frac{\exp(Q(s, a|\beta_j)\beta_j)}{\sum_{a' \in A_s} \exp(Q(s, a'|\beta_j)\beta_j)} \]

\[ \beta_j \sim \lnN(\mu, \sigma) \]
MDP as a Cognitive Model

What to do?

Markov Decision Process
MDP as a Cognitive Model

Beliefs

Goals

Transition Function

Reward Structure
MDP as a Cognitive Model

MDP-MM Parameter Space

\[ p\left( a \mid s, \beta, R, T \right) \]

- **R**: Goals / Motivation
- **T**: Beliefs about the world
- **β**: Decision Making
- **Actions**
MDP-MM Parameter Space

\[ p(a | s, \beta, R, T) \]

- **R**
  - Goals / Motivation
- **T**
  - Beliefs about the world
- **\( \beta \)**
  - Decision Making
- **Actions**

**Parameter Treatment**

<table>
<thead>
<tr>
<th>Parameter Treatment</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>{5, 1, ...}</td>
</tr>
<tr>
<td>Population</td>
<td>( R )</td>
</tr>
<tr>
<td>Group</td>
<td>( R_g )</td>
</tr>
<tr>
<td>Person</td>
<td>( R_j )</td>
</tr>
</tbody>
</table>
# MDP-MM Parameter Space

## Transition Parameters

### Totally Free

| A1  | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>$\lambda_{11}$</td>
<td>$\lambda_{12}$</td>
<td>$\lambda_{13}$</td>
<td>$\lambda_{14}$</td>
<td>$\lambda_{15}$</td>
<td>$\lambda_{16}$</td>
<td>$\lambda_{17}$</td>
<td>$\lambda_{18}$</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>$\lambda_{21}$</td>
<td>$\lambda_{22}$</td>
<td>$\lambda_{23}$</td>
<td>$\lambda_{24}$</td>
<td>$\lambda_{25}$</td>
<td>$\lambda_{26}$</td>
<td>$\lambda_{27}$</td>
<td>$\lambda_{28}$</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>$\lambda_{31}$</td>
<td>$\lambda_{32}$</td>
<td>$\lambda_{33}$</td>
<td>$\lambda_{34}$</td>
<td>$\lambda_{35}$</td>
<td>$\lambda_{36}$</td>
<td>$\lambda_{37}$</td>
<td>$\lambda_{38}$</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>$\lambda_{41}$</td>
<td>$\lambda_{42}$</td>
<td>$\lambda_{43}$</td>
<td>$\lambda_{44}$</td>
<td>$\lambda_{45}$</td>
<td>$\lambda_{46}$</td>
<td>$\lambda_{47}$</td>
<td>$\lambda_{48}$</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>$\lambda_{51}$</td>
<td>$\lambda_{52}$</td>
<td>$\lambda_{53}$</td>
<td>$\lambda_{54}$</td>
<td>$\lambda_{55}$</td>
<td>$\lambda_{56}$</td>
<td>$\lambda_{57}$</td>
<td>$\lambda_{58}$</td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>$\lambda_{61}$</td>
<td>$\lambda_{62}$</td>
<td>$\lambda_{63}$</td>
<td>$\lambda_{64}$</td>
<td>$\lambda_{65}$</td>
<td>$\lambda_{66}$</td>
<td>$\lambda_{67}$</td>
<td>$\lambda_{68}$</td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>$\lambda_{71}$</td>
<td>$\lambda_{72}$</td>
<td>$\lambda_{73}$</td>
<td>$\lambda_{74}$</td>
<td>$\lambda_{75}$</td>
<td>$\lambda_{76}$</td>
<td>$\lambda_{77}$</td>
<td>$\lambda_{78}$</td>
<td></td>
</tr>
</tbody>
</table>
# MDP-MM Parameter Space

## Transition Parameters

**Fixed by Objective Reality**

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>S5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2015 by Educational Testing Service. All rights reserved. ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). MEASURING THE POWER OF LEARNING is a trademark of ETS. 30141
### MDP-MM Parameter Space

**Transition Parameters**

<table>
<thead>
<tr>
<th>A1</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\lambda_1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1-$\lambda_1$</td>
<td>1-$\lambda_1$</td>
</tr>
<tr>
<td>S5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
MDP-MM Parameter Space

Transition Parameters

**Fixed by Misconception**

| A1 | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | ...
|----|----|----|----|----|----|----|----|----|------
| S1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0    |
| S2 | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0    |
| S3 | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0    |
| S4 | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0    |
| S5 | 0  | 0  | 0  | 0  | 0  | 0.1| 0  | 0  | 0.9  |
| S6 | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0    |
| S7 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0    |
MDP-MM Parameter Space

Transition Parameters

Categorical by Belief:

\[ T = \{H_1, H_2\} \]

- \( H_1 \rightarrow A1 \) may work in S4
- \( H_2 \rightarrow A1 \) may work in S5
MDP-MM Estimation

We use marginal maximum likelihood (MML) to estimate the population and group level parameters

\[
L(\xi|O) = \prod_{j=1}^{N} \prod_{t=1}^{T_j} p(a_{jt}|s_{jt}, \xi)
\]

\[
L(\mu, \sigma|O) = \int \prod_{t=1}^{T_j} \frac{\exp(Q(s_t, a_t|\beta_j)\beta_j)}{\sum_{a' \in A} \exp(Q(s_t, a'|\beta_j)\beta_j)} p(\beta_j|\mu, \sigma^2) d\beta_j,
\]
\[
\beta_j \sim \lnN(\mu, \sigma^2)
\]

And MLE to estimate the person level parameters.
MDP-MM Estimation

Q-Function is recursive – must be solved using dynamic programming.

\[
Q(s, a) = \sum_{s' \in S} p(s'|s, a) \left( R(s, a, s') + \gamma \sum_{a' \in A} p(a'|s') Q(s', a') \right)
\]
## Peg Solitaire Simulation Studies

Game boards with varying complexity

<table>
<thead>
<tr>
<th>Board</th>
<th>Path Length</th>
<th>Move Actions</th>
<th>Reachable States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny Cross</td>
<td>5</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>Big Cross</td>
<td>8</td>
<td>22</td>
<td>153</td>
</tr>
<tr>
<td>Big-L</td>
<td>13</td>
<td>30</td>
<td>807</td>
</tr>
<tr>
<td>Diamond</td>
<td>11</td>
<td>70</td>
<td>5923</td>
</tr>
</tbody>
</table>

- **Tiny Cross**
- **Big Cross**
- **Big L**
- **Diamond**
Peg Solitaire Parameters

No Transition Parameters.

Capability parameters: $\beta_j, \mu, \sigma$

Rewards:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function</th>
<th>Example Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{win}$</td>
<td>Reward for scoring with one peg left</td>
<td>5.0 Fixed</td>
</tr>
<tr>
<td>$R_{peg}$</td>
<td>Add to reward for each extra peg</td>
<td>-1.0 Fixed</td>
</tr>
<tr>
<td>$R_{move}$</td>
<td>Cost of a move</td>
<td>-0.1 Est.</td>
</tr>
<tr>
<td>$R_{reset}$</td>
<td>Cost of reset</td>
<td>-1.0 Est.</td>
</tr>
</tbody>
</table>
## Estimating Capability

### Graphs

- **Tiny Cross**
- **Big Cross**
- **Big L**
- **Diamond**

<table>
<thead>
<tr>
<th>Board</th>
<th>Ceiling Thresh.</th>
<th>Students Remaining</th>
<th>$\beta_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny Cross</td>
<td>2.03</td>
<td>0.80</td>
<td>-0.064</td>
</tr>
<tr>
<td>Big Cross</td>
<td>2.33</td>
<td>0.84</td>
<td>-0.036</td>
</tr>
<tr>
<td>Big-L</td>
<td>2.62</td>
<td>0.88</td>
<td>-0.072</td>
</tr>
<tr>
<td>Diamond</td>
<td>2.28</td>
<td>0.84</td>
<td>-0.045</td>
</tr>
</tbody>
</table>
Estimating Capability & Motivation

At the population level.
200 students/group. 25 games/student/board.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Capability</th>
<th>Motivation</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$R_{move}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>0.5</td>
<td>0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Low</td>
<td>0.5</td>
<td>0.75</td>
<td>-0.75</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>High</td>
<td>-0.5</td>
<td>0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>Low</td>
<td>-0.5</td>
<td>0.75</td>
<td>-0.75</td>
</tr>
</tbody>
</table>
Estimating Capability & Motivation

At the population level.
200 students/group. 25 games/student/board.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Capability</th>
<th>Motivation</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$R_{move}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>0.5</td>
<td>0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Low</td>
<td>0.5</td>
<td>0.75</td>
<td>-0.75</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>High</td>
<td>-0.5</td>
<td>0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>Low</td>
<td>-0.5</td>
<td>0.75</td>
<td>-0.75</td>
</tr>
</tbody>
</table>
## Estimating Capability & Motivation

At the population level.
200 students/group. 25 games/student/board.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Capability</th>
<th>Motivation</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$R_{move}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>High</td>
<td>0.5</td>
<td>0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Low</td>
<td>0.5</td>
<td>0.75</td>
<td>-0.75</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>High</td>
<td>-0.5</td>
<td>0.75</td>
<td>-0.05</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>Low</td>
<td>-0.5</td>
<td>0.75</td>
<td>-0.75</td>
</tr>
</tbody>
</table>
Estimating Capability & Motivation

[Box plots showing Mean Total Score by HAHM, HALM, LAHM, LALM categories]

[Scatter plot showing Mean Est. Ability vs. Mean Est. Move Penalty with different colored markers for Ability-Motivation categories: High-High, High-Low, Low-High, Low-Low]
MDP-MM in Action

PBS-Kids Microbes
Application: Microbes

N = 238 (148)
Application: Microbes

N = 238 (148)
# MDP Model for Microbes

6 Game Levels. Each modeled as a separate MDP

<table>
<thead>
<tr>
<th>State Space</th>
<th>State Variables:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Microbe Config</td>
</tr>
<tr>
<td></td>
<td>• Win History</td>
</tr>
<tr>
<td></td>
<td>= 484 States</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action Set</th>
<th>Buy Mito, Buy Chloro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Play Level, Stop</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rewards</th>
<th>Win, Lose, Buy</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Transitions</th>
<th>Play</th>
</tr>
</thead>
</table>
|             | => \[
|             | \begin{cases}
|             | \text{win} & p(\text{win}|s, a = \text{play}) \\
|             | \text{lose} & 1 - p(\text{win}|s, a = \text{play}) \\
|             | \end{cases} |
|             |      |
|             |      |
Estimating Capability

- Transition parameters are fixed.
- Rewards either fixed or estimated at the population level.

<table>
<thead>
<tr>
<th></th>
<th>Post-test Correlations</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDP-MM Fixed R</td>
<td>0.507</td>
<td>15465</td>
</tr>
<tr>
<td>MDP-MM Est R</td>
<td>0.516</td>
<td>11243</td>
</tr>
<tr>
<td>IRT First Try</td>
<td>0.317</td>
<td></td>
</tr>
<tr>
<td>IRT Multi-try PC</td>
<td>0.379</td>
<td></td>
</tr>
</tbody>
</table>

The estimates for $\beta_j$ from the MDP models correlated better with the posttest than the IRT estimates for $\theta_j$. 
Microbes Transition Parameters

\[
\text{Play} \rightarrow \begin{cases} 
\text{win} & p(\text{win}|s, a = \text{play}) \\
\text{lose} & 1 - p(\text{win}|s, a = \text{play}) 
\end{cases}
\]

To get at student beliefs, assume each student has an ideal microbe configuration.

\[c_j = \text{student } j\text{'s ideal } \# \text{ of chloroplasts}\]
\[m_j = \text{student } j\text{'s ideal } \# \text{ of mitochondria}\]

\[\max(p(\text{win}|s, a = \text{play})) = p(\text{win}|s = \{c_j, m_j\}, a = \text{play})\]
Estimating Beliefs/Understanding

Median posttest ability estimates:

-0.18  -0.36  -0.54  0.33
MDP-MM in Action

SimCityEDU Pollution Challenge
Application: SimCityEDU

N = 224
SimCity EDU

Sandbox style game
Many available actions
Ongoing simulated activity
Mission 4: Pollution Problems

Welcome to Sierra Madre!

This city has a problem with high air pollution. Find out how to lower pollution and keep the power level optimal.

You have 10 minutes. Can you avoid power blackouts?
SimCity Assessment

• Designed to assess Systems Thinking

• Students must optimize two variables simultaneously
SimCity MDP

**Action set is huge**
- Follow Sim named Joe Smith
- View Apartment Building Status
- Upgrade Garbage Dump
- Build Large Solar Power Plant
- Build Small Solar Power Plant
- Dezone Commercial (23,45)
- Zone as Residential (302,82)
- Bulldoze the Smith’s House
- Turn off Coal Plant
- Expand School
- Build Statue at City Hall ...

**State space is huge**
Game State includes
- Location and status of
  - Every Sim
  - Every Building
- Time of day
- City funds ($$)
- Severity and location of Pollution
- Wind direction and speed ...

Need to trim down to important subset!
# SimCity Actions

<table>
<thead>
<tr>
<th>Build</th>
<th>Turn On</th>
<th>Turn Off</th>
<th>Upgrade</th>
<th>Bulldoze</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Wind Wind</td>
<td>Wind Wind Wind</td>
<td>Wind Wind Wind</td>
<td>Wind Wind Wind</td>
<td>Wind Wind Wind</td>
</tr>
<tr>
<td>Coal Coal Coal</td>
<td>Coal Coal Coal</td>
<td>Coal Coal Coal</td>
<td>Coal Coal Coal</td>
<td>Coal Coal Coal</td>
</tr>
</tbody>
</table>

Wind + Solar + Coal + Wait + End Mission = 17 actions
# SimCity State Space

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th># Values</th>
</tr>
</thead>
<tbody>
<tr>
<td># Coal Generators On</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td># Coal Generators Off</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td># Wind Turbines</td>
<td>0</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td># Solar Panels</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Power Balance</td>
<td>-8</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>Pollution</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Cash</td>
<td>0</td>
<td>30</td>
<td>31</td>
</tr>
</tbody>
</table>

Total # of States: 2,856,960
But only 25,420 reachable states
SimCity Rewards

**MISSION OBJECTIVE**

- AQI below 100 and no blackout

**BONUS OBJECTIVES**

- AQI below 50
- Power was never dangerously low

The air quality index (AQI) in the city was 59. The power capacity was 26.9 MW. The power needed was 23.3 MW.

Good effort! Swap your coal plants to lower air pollution even more. Just be careful not to cause a power failure!

**YOU EARNED A BRONZE MEDAL!**

What's Next?

- PLAY AGAIN
- ALL DONE!

NEXT MISSION
SimCity Rewards

**MISSION OBJECTIVE**

- AQI below 100 and no blackout

**BONUS OBJECTIVES**

- AQI below 50
- Power was never dangerously low

The air quality index (AQI) in the city was 83. The power capacity was 41.3 MW. The power needed was 23.3 MW.

The air could be cleaner. The good news is that you didn't have a temporary power failure. Have you opened the pollution map?

YOU EARNED A SILVER MEDAL!

What's Next?

PLAY AGAIN  ALL DONE!

NEXT MISSION
SimCity Rewards

**MISSION OBJECTIVE**

- AQI below 100 and no blackout

**BONUS OBJECTIVES**

- AQI below 50
- Power was never dangerously low

The air quality index (AQI) in the city was 43. The power capacity was 32.8 MW. The power needed was 23.3 MW.

Wow, you're a great mayor! You kept the power optimal and reduced air pollution. Can you teach your friends how to be such an awesome mayor?
## SimCity Rewards

<table>
<thead>
<tr>
<th>Goal Hypotheses</th>
<th>Bronze Medal</th>
<th>Pollution Silver</th>
<th>Power Silver</th>
<th>Gold Medal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medals</td>
<td>+5</td>
<td>+5</td>
<td>+5</td>
<td>+10</td>
</tr>
<tr>
<td>Just Win</td>
<td>+10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pollution</td>
<td>+5</td>
<td>+5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Power</td>
<td>+5</td>
<td>0</td>
<td>+5</td>
<td>0</td>
</tr>
</tbody>
</table>
# Estimating Goals

<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood</th>
<th>Num Students Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medals</td>
<td>-17565.1</td>
<td>38</td>
</tr>
<tr>
<td>Just Win</td>
<td>-17974.5</td>
<td>28</td>
</tr>
<tr>
<td>Pollution</td>
<td>-17974.6</td>
<td>38</td>
</tr>
<tr>
<td>Power</td>
<td>-17915.0</td>
<td>24</td>
</tr>
</tbody>
</table>
## Estimating Goals

<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood</th>
<th>Num Students Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medals</td>
<td>-17565.1</td>
<td>38</td>
</tr>
<tr>
<td>Just Win</td>
<td>-17974.5</td>
<td>28</td>
</tr>
<tr>
<td>Pollution</td>
<td>-17974.6</td>
<td>38</td>
</tr>
<tr>
<td>Power</td>
<td>-17915.0</td>
<td>24</td>
</tr>
</tbody>
</table>

53 Students who fit none – Posteriors were flat
Overall Sufficiency of SC MDP

Strong implication that our model is over simplified for many of the students.

Consider expanding model:
- Zoning actions?
- Sim Happiness as a goal?
Conclusions

Markov Decision Process Measurement Model

- Potential as a flexible framework for assessment
  - Estimate general ability from task process data
  - Separate student motivation, system understanding and strategic ability
- Sensitive to specification of cognitive processes
  - Microbe data truncated after first win, not sure what the student was motivated by at that point.
Conclusions

• Early work; much yet to do
  • Improve algorithms & estimation
  • Gather more validity evidence
  • Partially Observable MDP (POMDP)

Just one example of Cognitive Process Models for assessment
Other work:

• **Multi-modal analytics: evidence from stream data**
  • Emotion detection
  • Gestures, posture, and actions
  • Voice tone and fluency

• **Assessing Collaboration**
  • Collaborative Assessment Frame
  • Collaborative Dialog Analysis
  • Social Network Models
  • Hawkes Process Models
This work was made possible by ...

**Colleagues**
Anna Rafferty
Tom Griffiths
Sophia Rabe-Hesketh
Mark Wilson
Matt Silverglitt
Malcolm Bauer
Alina von Davier

**Data Provided By**
WestEd

GlassLab

RED HILL STUDIOS
Using Markov Decision Processes to Understand Student Thinking in Performance Tasks

Michelle M. LaMar
Educational Testing Service
mlamar@ets.org

October 1st, 2015
References


## Simulation Study 1

Results for recovery of population parameters

<table>
<thead>
<tr>
<th>Board</th>
<th>$\mu$ Bias</th>
<th>$\mu$ RMSE</th>
<th>$\sigma$ Bias</th>
<th>$\sigma$ RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny Cross</td>
<td>0.017</td>
<td>0.055</td>
<td>0.103</td>
<td>0.217</td>
</tr>
<tr>
<td>Big Cross</td>
<td>0.028</td>
<td>0.053</td>
<td>0.076</td>
<td>0.121</td>
</tr>
<tr>
<td>Big-L</td>
<td>0.019</td>
<td>0.058</td>
<td>0.056</td>
<td>0.093</td>
</tr>
<tr>
<td>Diamond</td>
<td>0.012</td>
<td>0.058</td>
<td>0.056</td>
<td>0.105</td>
</tr>
</tbody>
</table>
Simulation Study 1

Results for recovery of population parameters

<table>
<thead>
<tr>
<th>Board</th>
<th>$\mu$ Bias</th>
<th>$\mu$ RMSE</th>
<th>$\sigma$ Bias</th>
<th>$\sigma$ RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny Cross</td>
<td>0.017 (0.008)</td>
<td>0.055 (0.023)</td>
<td>0.103 (0.029)</td>
<td>0.217 (0.102)</td>
</tr>
<tr>
<td>Big Cross</td>
<td>0.028 (0.007)</td>
<td>0.053 (0.024)</td>
<td>0.076 (0.014)</td>
<td>0.121 (0.053)</td>
</tr>
<tr>
<td>Big-L</td>
<td>0.019 (0.008)</td>
<td>0.058 (0.034)</td>
<td>0.056 (0.011)</td>
<td>0.093 (0.049)</td>
</tr>
<tr>
<td>Diamond</td>
<td>0.012 (0.009)</td>
<td>0.058 (0.033)</td>
<td>0.056 (0.013)</td>
<td>0.105 (0.046)</td>
</tr>
</tbody>
</table>
Simulation Study 1

Results for recovery of $\beta_j$

<table>
<thead>
<tr>
<th>Board</th>
<th>Full Population Bias</th>
<th>Full Population RMSE</th>
<th>Ceiling Thresh.</th>
<th>Students Remaining</th>
<th>Truncated Pop. Bias</th>
<th>Truncated Pop. RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny Cross</td>
<td>-0.318</td>
<td>0.738</td>
<td>2.03</td>
<td>0.80</td>
<td>-0.064</td>
<td>0.395</td>
</tr>
<tr>
<td>Big Cross</td>
<td>-0.228</td>
<td>0.631</td>
<td>2.33</td>
<td>0.84</td>
<td>-0.036</td>
<td>0.362</td>
</tr>
<tr>
<td>Big-L</td>
<td>-0.208</td>
<td>0.571</td>
<td>2.62</td>
<td>0.88</td>
<td>-0.072</td>
<td>0.365</td>
</tr>
<tr>
<td>Diamond</td>
<td>-0.220</td>
<td>0.587</td>
<td>2.28</td>
<td>0.84</td>
<td>-0.045</td>
<td>0.327</td>
</tr>
</tbody>
</table>

Return
MDP-MM Properties

Decision Difficulty

$\Delta Q(s, a'|\beta) = Q(s, \pi(s)) - Q(s, a'|\beta)$
MDP-MM Properties

Decision Difficulty

$$\Delta Q(s, a' | \beta) = Q(s, \pi(s)) - Q(s, a' | \beta)$$
Application: Microbes

N = 238 (148)
Educational Game: Microbes
Educational Game: Microbes
Educational Game: Microbes

Needs energy from food or sunlight
Mitochondrion Needs energy from food or sunlight
Educational Game: Microbes

Needs energy from food or sunlight
Educational Game: Microbes

This water drop has nothing but water, carbon dioxide, and light. You’ll need to make your own food. Check out the new options at the Mart.

<table>
<thead>
<tr>
<th>Locomotion</th>
<th>Micro Engines</th>
<th>Food Generators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These guys make light work of getting food - they make it themselves little carbon dioxide... you know, photosynthesis?

- **Chloroplasts** (unit cost 5) | 5
- **Chloroplast (1)**
- **Chloroplast (3)**
- **Chloroplasts (unit cost 5)** | 0

**OBJECTIVE**

- **Task 1**: Find the pockets of light so that your microorganism can make food. But watch out - the predators tend to hang out there:
  - Earn 10 tokens

---

Copyright © 2015 by Educational Testing Service. All rights reserved. ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). MEASURING THE POWER OF LEARNING is a trademark of ETS. 30141
Educational Game: Microbes
Transition Function for Microbes

Model the probability of play success as a logistic regression:

$$logit(p(\text{success}|\text{play}, s)) =$$

$$\beta_0 + \beta_1 f_l + \beta_2 f_l m_s + \beta_3 s_l + \beta_4 s_l c_s + \beta_5 s_l r_s$$

$$f_l = \text{food available (0,1,2,3)}$$

$$s_l = \text{sunlight (0,1,2,3)}$$

$$m_s = \text{number mitochondria}$$

$$c_s = \text{number chloroplasts}$$
Transition Function for Microbes

The factors are a combination of environmental, based on the game level

\[ \beta_0 + \beta_1 f_l + \beta_2 f_l m_s + \beta_3 s_l + \beta_4 s_l c_s + \beta_5 s_l r_s \]

- \( f_l = \) food available (0, 1, 2, 3)
- \( s_l = \) sunlight (0, 1, 2, 3)
- \( m_s = \) number mitochondria
- \( c_s = \) number chloroplasts
Transition Function for Microbes

The factors are a combination of environmental, based on the game level and the microbe configuration:

\[
\beta_0 + \beta_1 f_l + \beta_2 f_l m_s + \beta_3 s_l + \beta_4 s_l c_s + \beta_5 s_l r_s
\]

- \( f_l \) = food available (0,1,2,3)
- \( s_l \) = sunlight (0,1,2,3)
- \( m_s \) = number mitochondria
- \( c_s \) = number chloroplasts