

Survey-based vs. Incentivized Experimental Measures

Teodora Boneva (UCL/HCEO)

October 2015

Overview

1 Advantages and Disadvantages

- Controlling the Environment
- Feasibility
- “Faking”
- Availability

2 Experimental Validation of Survey Methods

- Falk et al. (2014)
- Other Examples

Controlling the Environment

- **Incentivized Experimental Measures:**
 - ▶ observe choices in controlled environment
- Surveys:
 - ▶ we do not know how subjects interpret the questions
 - ▶ we do not have information on the environment subjects face/how they perceive their environment
 - ▶ we do not know what reference points/norms subjects use

Controlling the Environment

- Incentivized Experimental Measures:
 - ▶ observe choices in controlled environment
- Surveys:
 - ▶ we do not know how subjects interpret the questions
 - ▶ we do not have information on the environment subjects face/how they perceive their environment
 - ▶ we do not know what reference points/norms subjects use

Controlling the Environment

- Examples:

- ▶ *How willing are you to take risks in the context of car driving? (GSOEP, Dohmen et al., 2011)*
- ▶ *I see myself as someone who is curious about many different things. (Big 5, Openness to Experience, John and Srivastava, 1999)*
- ▶ *I have been obsessed with a certain idea or project for a short time but later lost interest. (Short Grit Scale, Duckworth and Quinn, 2009)*

Controlling the Environment

- Examples:

- ▶ *How willing are you to take risks in the context of car driving? (GSOEP, Dohmen et al., 2011)*
- ▶ *I see myself as someone who is curious about many different things. (Big 5, Openness to Experience, John and Srivastava, 1999)*
- ▶ *I have been obsessed with a certain idea or project for a short time but later lost interest. (Short Grit Scale, Duckworth and Quinn, 2009)*

Controlling the Environment

- Examples:

- ▶ *How willing are you to take risks in the context of car driving?*
(GSOEP, Dohmen et al., 2011)
- ▶ *I see myself as someone who is curious about many different things.*
(Big 5, Openness to Experience, John and Srivastava, 1999)
- ▶ *I have been obsessed with a certain idea or project for a short time but later lost interest.* (Short Grit Scale, Duckworth and Quinn, 2009)

Feasibility

- **Incentivized Experimental Measures:**
 - ▶ expensive
 - ▶ difficult to administer: experimenter time, payments, interactive games
- Surveys:
 - ▶ cheap
 - ▶ easy to administer

Feasibility

- Incentivized Experimental Measures:
 - ▶ expensive
 - ▶ difficult to administer: experimenter time, payments, interactive games
- Surveys:
 - ▶ cheap
 - ▶ easy to administer

“Faking”

- respondents might have an incentive to “fake” traits
 - ▶ impression management
 - ▶ self-deception
- how important these factors are might depend on
 - ▶ personal characteristics
 - ▶ context of the survey

“Faking”

- respondents might have an incentive to “fake” traits
 - ▶ impression management
 - ▶ self-deception
- how important these factors are might depend on
 - ▶ personal characteristics
 - ▶ context of the survey

“Faking”

- easy to see which qualities might be valuable
- *Example: I see myself as someone who*
 - ▶ *can be moody (neuroticism)*
 - ▶ *worries a lot (neuroticism)*
 - ▶ *can be somewhat careless (conscientiousness)*
 - ▶ *tends to be lazy (conscientiousness)*
- somewhat less problematic with incentivized experiments because real stakes are involved

“Faking”

- easy to see which qualities might be valuable
- *Example: I see myself as someone who*
 - ▶ *can be moody (neuroticism)*
 - ▶ *worries a lot (neuroticism)*
 - ▶ *can be somewhat careless (conscientiousness)*
 - ▶ *tends to be lazy (conscientiousness)*
- somewhat less problematic with incentivized experiments because real stakes are involved

“Faking”

- easy to see which qualities might be valuable
- *Example: I see myself as someone who*
 - ▶ *can be moody (neuroticism)*
 - ▶ *worries a lot (neuroticism)*
 - ▶ *can be somewhat careless (conscientiousness)*
 - ▶ *tends to be lazy (conscientiousness)*
- somewhat less problematic with incentivized experiments because real stakes are involved

Availability

- Preference parameters
 - ▶ time discounting
 - ▶ risk aversion
 - ▶ social preferences
- survey-based measures exist
- experiments exist

Availability

- Preference parameters
 - ▶ time discounting
 - ▶ risk aversion
 - ▶ social preferences
- survey-based measures exist
- experiments exist

Availability

- Time preferences:
 - ▶ Choice between sooner and later payments, Marshmallow task
 - ▶ *'How patient are you on a scale from 1 to 10?'* (GSOEP)
- Risk preferences:
 - ▶ Lottery choice tasks, Devil's Task, Balloon Analogue Risk Task
 - ▶ *'How willing are you to take risks in general on a scale from 1 to 10?'* (GSOEP)
- Reciprocity:
 - ▶ Ultimatum game, Gift exchange game
 - ▶ *'If someone does me a favor, I am prepared to return it.'* (GSOEP)
 - ▶ *'If I suffer a serious wrong, I will take revenge as soon as possible, irrespective of the cost.'* (GSOEP)

Availability

- Time preferences:
 - ▶ Choice between sooner and later payments, Marshmallow task
 - ▶ *'How patient are you on a scale from 1 to 10?'* (GSOEP)
- Risk preferences:
 - ▶ Lottery choice tasks, Devil's Task, Balloon Analogue Risk Task
 - ▶ *'How willing are you to take risks in general on a scale from 1 to 10?'* (GSOEP)
- Reciprocity:
 - ▶ Ultimatum game, Gift exchange game
 - ▶ *'If someone does me a favor, I am prepared to return it.'* (GSOEP)
 - ▶ *'If I suffer a serious wrong, I will take revenge as soon as possible, irrespective of the cost.'* (GSOEP)

Availability

- Time preferences:
 - ▶ Choice between sooner and later payments, Marshmallow task
 - ▶ *'How patient are you on a scale from 1 to 10?'* (GSOEP)
- Risk preferences:
 - ▶ Lottery choice tasks, Devil's Task, Balloon Analogue Risk Task
 - ▶ *'How willing are you to take risks in general on a scale from 1 to 10?'* (GSOEP)
- Reciprocity:
 - ▶ Ultimatum game, Gift exchange game
 - ▶ *'If someone does me a favor, I am prepared to return it.'* (GSOEP)
 - ▶ *'If I suffer a serious wrong, I will take revenge as soon as possible, irrespective of the cost.'* (GSOEP)

Availability

- **Personality measures**

- ▶ **Big 5:**

- Conscientiousness
 - Openness to new experience
 - Neuroticism
 - Extraversion
 - Agreeableness

- ▶ **Curiosity**

- ▶ **Grit**

- survey-based measures exist

- for many personality measures there are no analogous experiments

Availability

- Personality measures
 - ▶ Big 5:
 - Conscientiousness
 - Openness to new experience
 - Neuroticism
 - Extraversion
 - Agreeableness
 - ▶ Curiosity
 - ▶ Grit
- survey-based measures exist
- for many personality measures there are no analogous experiments

Availability

- Big 5
 - ▶ ???
- Curiosity
 - ▶ ???
- Grit:
 - ▶ Incentivized Grit Task measuring choice of task difficulty, perseverance after negative feedback, goal setting and skill accumulation (Alan et al., 2015)
 - ▶ Anagram Task (Gerhards and Gravert, 2015)

Availability

- Big 5
 - ▶ ???
- Curiosity
 - ▶ ???
- Grit:
 - ▶ Incentivized Grit Task measuring choice of task difficulty, perseverance after negative feedback, goal setting and skill accumulation (Alan et al., 2015)
 - ▶ Anagram Task (Gerhards and Gravert, 2015)

Availability

- Big 5
 - ▶ ???
- Curiosity
 - ▶ ???
- Grit:
 - ▶ **Incentivized Grit Task** measuring choice of task difficulty, perseverance after negative feedback, goal setting and skill accumulation (Alan et al., 2015)
 - ▶ **Anagram Task** (Gerhards and Gravert, 2015)

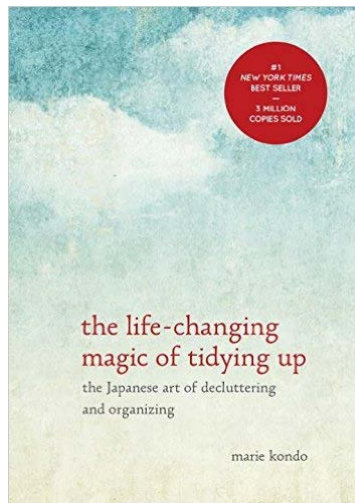
Availability

- Big 5
 - ▶ ???
- Curiosity
 - ▶ ???
- Grit:
 - ▶ **Incentivized Grit Task** measuring choice of task difficulty, perseverance after negative feedback, goal setting and skill accumulation (Alan et al., 2015)
 - ▶ **Anagram Task** (Gerhards and Gravert, 2015)

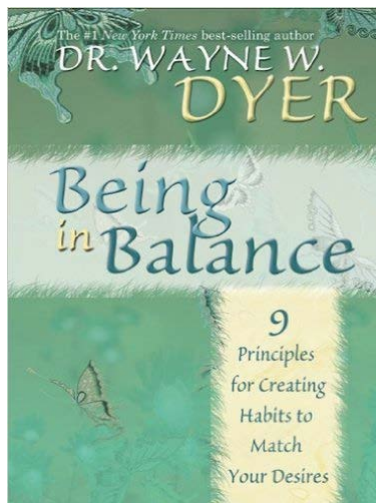
Beliefs

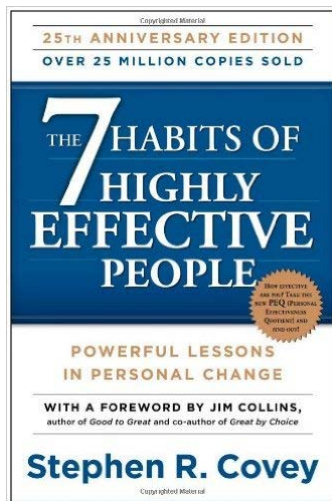
- Importance of beliefs for skill accumulation:
 - ▶ beliefs about productivity of effort/investments
 - ▶ beliefs about malleability of skills
 - ▶ beliefs about malleability of personality
- Can beliefs be seen as a 'skill'?
- Use of hypothetical scenarios to elicit beliefs

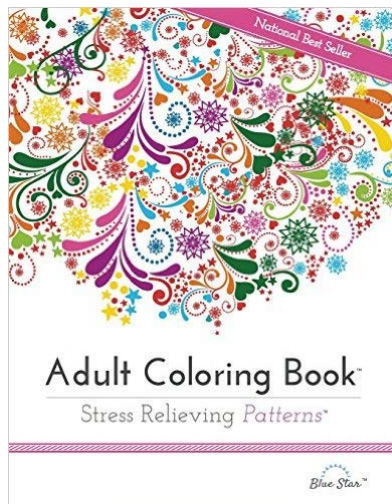
Beliefs



Beliefs







Overview

1 Advantages and Disadvantages

- Controlling the Environment
- Feasibility
- “Faking”
- Availability

2 Experimental Validation of Survey Methods

- Falk et al. (2014)
- Other Examples

Experimental Validation of Surveys - Falk et al. (2014)

- develops experimentally-validated survey modules of economic preferences:
 - ▶ risk preference
 - ▶ time preference
 - ▶ altruism
 - ▶ trust
 - ▶ positive reciprocity
 - ▶ negative reciprocity
- subjects participate in experiments and fill out surveys (N=409)

Experimental Validation

Incentivized experiments treated as “gold standard” and survey questions are selected so that they have the greatest predictive power for behavior in the experimental task.

Experimental Validation of Surveys - Falk et al. (2014)

- develops experimentally-validated survey modules of economic preferences:
 - ▶ risk preference
 - ▶ time preference
 - ▶ altruism
 - ▶ trust
 - ▶ positive reciprocity
 - ▶ negative reciprocity
- subjects participate in experiments and fill out surveys (N=409)

Experimental Validation

Incentivized experiments treated as “gold standard” and survey questions are selected so that they have the greatest predictive power for behavior in the experimental task.

Experimental Validation of Surveys - Falk et al. (2014)

- develops experimentally-validated survey modules of economic preferences:
 - ▶ risk preference
 - ▶ time preference
 - ▶ altruism
 - ▶ trust
 - ▶ positive reciprocity
 - ▶ negative reciprocity
- subjects participate in experiments and fill out surveys (N=409)

Experimental Validation

Incentivized experiments treated as “gold standard” and survey questions are selected so that they have the greatest predictive power for behavior in the experimental task.

Experimental Validation of Surveys - Falk et al. (2014)

- two incentivized experiments for each preference (measurement error)
- experiments and surveys conducted one week apart (desire to be consistent)
- order reversed for half the subjects
- questions that best predict behavior in experiments
 - ▶ quantitative question: hypothetical version of experiment
 - ▶ qualitative question: subjective assessment of general orientation
- explained variance
 - ▶ test-retest: R^2 of 0.33-0.66
 - ▶ surveys: R^2 of 0.15-0.47

Experimental Validation of Surveys - Falk et al. (2014)

- two incentivized experiments for each preference (measurement error)
- experiments and surveys conducted one week apart (desire to be consistent)
- order reversed for half the subjects
- questions that best predict behavior in experiments
 - ▶ quantitative question: hypothetical version of experiment
 - ▶ qualitative question: subjective assessment of general orientation
- explained variance
 - ▶ test-retest: R^2 of 0.33-0.66
 - ▶ surveys: R^2 of 0.15-0.47

Experimental Validation of Surveys - Falk et al. (2014)

- two incentivized experiments for each preference (measurement error)
- experiments and surveys conducted one week apart (desire to be consistent)
- order reversed for half the subjects
- questions that best predict behavior in experiments
 - ▶ quantitative question: hypothetical version of experiment
 - ▶ qualitative question: subjective assessment of general orientation
- explained variance
 - ▶ test-retest: R^2 of 0.33-0.66
 - ▶ surveys: R^2 of 0.15-0.47

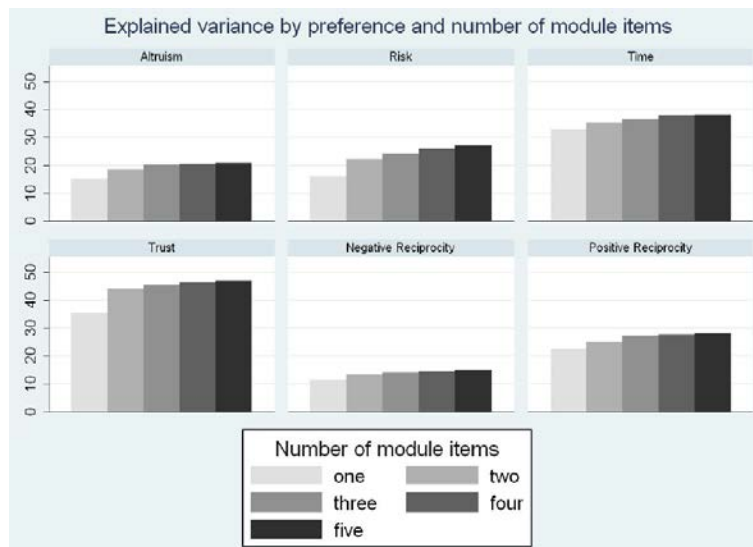
Experimental Validation of Surveys - Falk et al. (2014)

- two incentivized experiments for each preference (measurement error)
- experiments and surveys conducted one week apart (desire to be consistent)
- order reversed for half the subjects
- questions that best predict behavior in experiments
 - ▶ quantitative question: hypothetical version of experiment
 - ▶ qualitative question: subjective assessment of general orientation
- explained variance
 - ▶ test-retest: R^2 of 0.33-0.66
 - ▶ surveys: R^2 of 0.15-0.47

Experimental Validation of Surveys - Falk et al. (2014)

- two incentivized experiments for each preference (measurement error)
- experiments and surveys conducted one week apart (desire to be consistent)
- order reversed for half the subjects
- questions that best predict behavior in experiments
 - ▶ quantitative question: hypothetical version of experiment
 - ▶ qualitative question: subjective assessment of general orientation
- explained variance
 - ▶ test-retest: R^2 of 0.33-0.66
 - ▶ surveys: R^2 of 0.15-0.47

Experimental Validation of Surveys - Falk et al. (2014)



Experimental Validation of Surveys - Falk et al. (2014)

- Validation performed in
 - ▶ a non-representative sample
 - ▶ in one country
- structure of correlations between experimental measures and survey measures could differ
 - ▶ across individuals with different characteristics (e.g. IQ, age, gender, wealth)
 - ▶ across cultural contexts
- subjects with different characteristics/in different environments might interpret questions differently

Experimental Validation of Surveys - Falk et al. (2014)

- Validation performed in
 - ▶ a non-representative sample
 - ▶ in one country
- structure of correlations between experimental measures and survey measures could differ
 - ▶ across individuals with different characteristics (e.g. IQ, age, gender, wealth)
 - ▶ across cultural contexts
- subjects with different characteristics/in different environments might interpret questions differently

Experimental Validation of Surveys - Falk et al. (2014)

- Validation performed in
 - ▶ a non-representative sample
 - ▶ in one country
- structure of correlations between experimental measures and survey measures could differ
 - ▶ across individuals with different characteristics (e.g. IQ, age, gender, wealth)
 - ▶ across cultural contexts
- subjects with different characteristics/in different environments might interpret questions differently

Experimental Validation of Surveys - Other Examples

- Dohmen et al. (2011):
 - ▶ representative sample in Germany (N=450)
 - ▶ experimentally elicited risk attitudes correlate with 'willingness to take risk in general'
- Vieider et al. (2013):
 - ▶ ca. 3000 subjects in 30 countries (non-representative)
 - ▶ experimentally elicited risk/uncertainty attitudes and correlate with 'willingness to take risk in general'
 - ▶ size of correlation varies enormously across countries (-0.13 to +0.42)
- Vischer et al. (2013):
 - ▶ representative sample in Germany (N=839)
 - ▶ experimentally elicited time preferences correlate with 'tendency to be patient in general'
- Fehr et al. (2013):
 - ▶ representative sample in Germany (N=429)
 - ▶ experimentally elicited trust correlates with self-rated 'trusting behavior' and 'willingness to trust strangers'

Experimental Validation of Surveys - Other Examples

- Dohmen et al. (2011):
 - ▶ representative sample in Germany (N=450)
 - ▶ experimentally elicited risk attitudes correlate with 'willingness to take risk in general'
- Vieider et al. (2013):
 - ▶ ca. 3000 subjects in 30 countries (non-representative)
 - ▶ experimentally elicited risk/uncertainty attitudes and correlate with 'willingness to take risk in general'
 - ▶ size of correlation varies enormously across countries (-0.13 to +0.42)
- Vischer et al. (2013):
 - ▶ representative sample in Germany (N=839)
 - ▶ experimentally elicited time preferences correlate with 'tendency to be patient in general'
- Fehr et al. (2013):
 - ▶ representative sample in Germany (N=429)
 - ▶ experimentally elicited trust correlates with self-rated 'trusting behavior' and 'willingness to trust strangers'

Experimental Validation of Surveys - Other Examples

- Dohmen et al. (2011):
 - ▶ representative sample in Germany (N=450)
 - ▶ experimentally elicited risk attitudes correlate with 'willingness to take risk in general'
- Vieider et al. (2013):
 - ▶ ca. 3000 subjects in 30 countries (non-representative)
 - ▶ experimentally elicited risk/uncertainty attitudes and correlate with 'willingness to take risk in general'
 - ▶ size of correlation varies enormously across countries (-0.13 to +0.42)
- Vischer et al. (2013):
 - ▶ representative sample in Germany (N=839)
 - ▶ experimentally elicited time preferences correlate with 'tendency to be patient in general'
- Fehr et al. (2013):
 - ▶ representative sample in Germany (N=429)
 - ▶ experimentally elicited trust correlates with self-rated 'trusting behavior' and 'willingness to trust strangers'

Experimental Validation of Surveys - Other Examples

- Dohmen et al. (2011):
 - ▶ representative sample in Germany (N=450)
 - ▶ experimentally elicited risk attitudes correlate with 'willingness to take risk in general'
- Vieider et al. (2013):
 - ▶ ca. 3000 subjects in 30 countries (non-representative)
 - ▶ experimentally elicited risk/uncertainty attitudes and correlate with 'willingness to take risk in general'
 - ▶ size of correlation varies enormously across countries (-0.13 to +0.42)
- Vischer et al. (2013):
 - ▶ representative sample in Germany (N=839)
 - ▶ experimentally elicited time preferences correlate with 'tendency to be patient in general'
- Fehr et al. (2013):
 - ▶ representative sample in Germany (N=429)
 - ▶ experimentally elicited trust correlates with self-rated 'trusting behavior' and 'willingness to trust strangers'

Experimental Validation of Surveys - Other Examples

- Dohmen et al. (2011):
 - ▶ representative sample in Germany (N=450)
 - ▶ experimentally elicited risk attitudes correlate with 'willingness to take risk in general'
- Vieider et al. (2013):
 - ▶ ca. 3000 subjects in 30 countries (non-representative)
 - ▶ experimentally elicited risk/uncertainty attitudes and correlate with 'willingness to take risk in general'
 - ▶ size of correlation varies enormously across countries (-0.13 to +0.42)
- Vischer et al. (2013):
 - ▶ representative sample in Germany (N=839)
 - ▶ experimentally elicited time preferences correlate with 'tendency to be patient in general'
- Fehr et al. (2013):
 - ▶ representative sample in Germany (N=429)
 - ▶ experimentally elicited trust correlates with self-rated 'trusting behavior' and 'willingness to trust strangers'

Avenues for Research

- design incentivized experimental measures of non-cognitive skills
- develop experimentally-validated surveys
- gain better understanding of how survey-based measures and experimental measures correlate
 - ▶ for people with different characteristics
 - ▶ for people in different cultures
- gain better understanding of which measure is measuring what and how to decide between which measure(s) we want to use

Thank you!