

COGS 230: Human-Computer Interaction

Methodology Matters: How to do Experiments?

Julian Skirzyński

October 7, 2024

Authors



David Martin

Head of Psychology at NCSU:
Engineering Psychology, Research
Methods, Human Performance



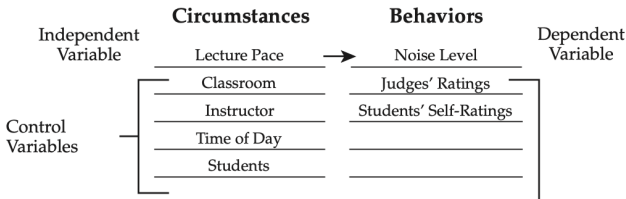
Joseph McGrath

Head of Psychology at UIUC:
Research Methods, Technology in
Groups, Social Psychology

Topic

"Doing research" simply means the systematic use of some set of theoretical and empirical tools to try to increase our understanding of some set of phenomena or events

Experimental Setup



- ▶ Independent variables
 - ▶ Variables independent of human behavior that we vary
 - ▶ Changing the value of the independent variable = "intervention"
- ▶ Dependent variables
 - ▶ Variables that quantify human behavior, i.e. our measures
- ▶ Control variables
 - ▶ Variables dependent on human behavior that we control for and keep constant

Question

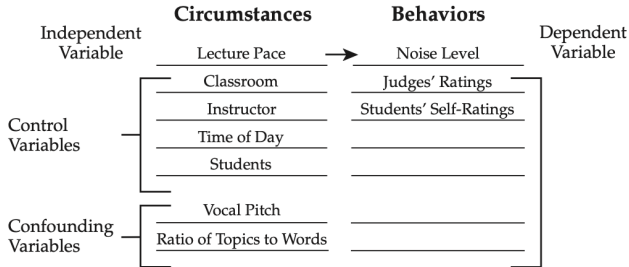
Why do we control for other variables?

Question

Why do we control for other variables?

To know that the change we are measuring stems from the change in the dependent variable

Experimental Setup (cont.)



- ▶ Independent variables
- ▶ Dependent variables
- ▶ Control variables
- ▶ Confounding variables
 - ▶ Variables changing systematically with the dependent variable
 - ▶ e.g., younger participants in one condition, older in the other

Question

Can we control for all confounding variables?

Question

Can we control for all confounding variables?

No

Question

Can we control for all confounding variables?

No

Why not?

Question

Can we control for all confounding variables?

No

Why not?

There may be just too many, and often there is

Question

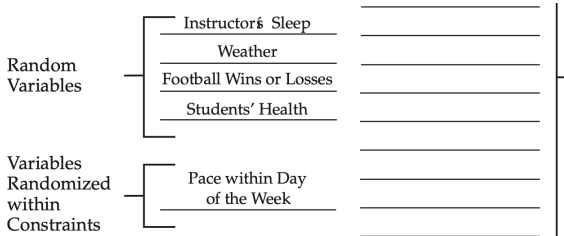
What can we do then?

Question

What can we do then?

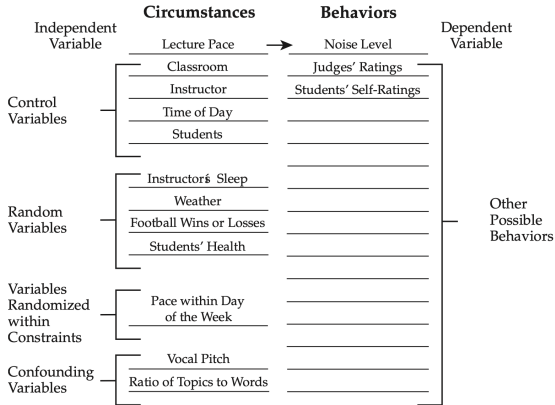
Use randomization

Experimental Setup (cont.)



- ▶ Random and semi-random variables
 - ▶ Variables which we want to vary across the levels of the independent variables and expect to take on random values
 - ▶ Some variables are semi-random or "random with constraints", i.e. are controlled in some aspects

Hypothesis



- ▶ A statement expressing a relationship between independent variables and the dependent variable(s)
- ▶ Direction of the hypothesis: expression controlling the direction of change of the dependent variable (e.g. increase/decrease)

Constructing Credible Experiments

- ▶ Use multiple methods to test our hypotheses

Question

What does using multiple methods give us?

Question

What does using multiple methods give us?

**Consistent results improve the credibility of the result,
whereas inconsistent result let us qualify our research**

Question

What does using multiple methods give us?

Consistent results improve the credibility of the result, whereas inconsistent result let us qualify our research

How does using multiple methods manifest in science?

Question

What does using multiple methods give us?

**Consistent results improve the credibility of the result,
whereas inconsistent result let us qualify our research**

How does using multiple methods manifest in science?

E.g. replication studies

Constructing Credible Experiments

- ▶ Use multiple methods to test our hypotheses
- ▶ Determine the most important factors to increase validity:
 - ▶ Generalizability through actors

Constructing Credible Experiments

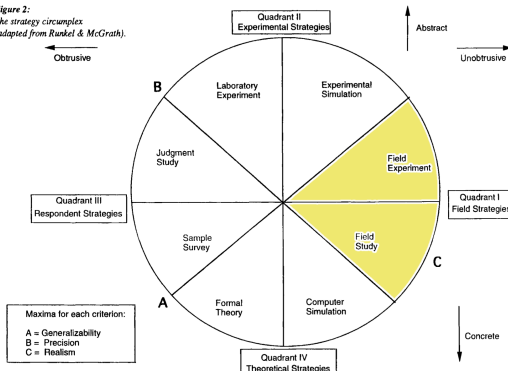
- ▶ Use multiple methods to test our hypotheses
- ▶ Determine the most important factors to increase validity:
 - ▶ Generalizability through actors
 - ▶ Precision of measuring the behavior

Constructing Credible Experiments

- ▶ Use multiple methods to test our hypotheses
- ▶ Determine the most important factors to increase validity:
 - ▶ Generalizability through actors
 - ▶ Precision of measuring the behavior
 - ▶ Realism of the study

Research Strategies: Setting is important

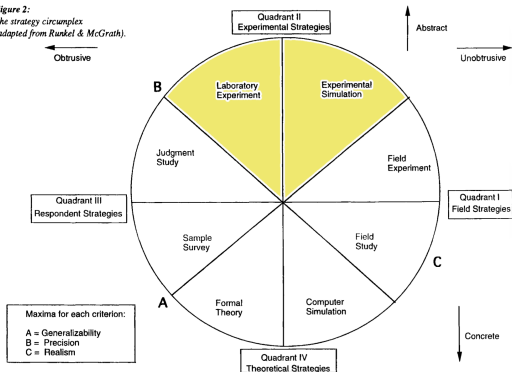
Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



- ▶ High realism, low precision, lower generalizability
- ▶ Field study: measure the dependent variable in the natural environment
- ▶ Field experiment: field study where you manipulate one major independent variable

Research Strategies: Measurement is more important

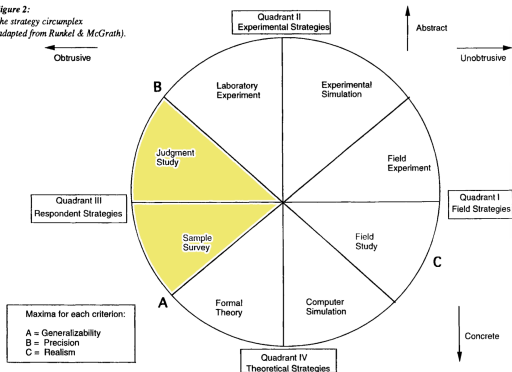
Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



- ▶ Lower realism, high precision, higher generalizability
- ▶ Lab experiment: measure the dependent variable in an artificial task
- ▶ Simulation experiment: measure the dependent variable in a task simulating the real world

Research Strategies: Setting does not matter

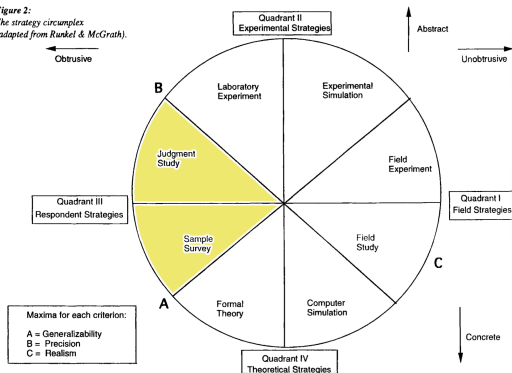
Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



- ▶ No realism, low precision, higher generalizability
- ▶ Sample survey: measure the dependent variable by administering a survey to a random batch of participants

Research Strategies: Setting does not matter

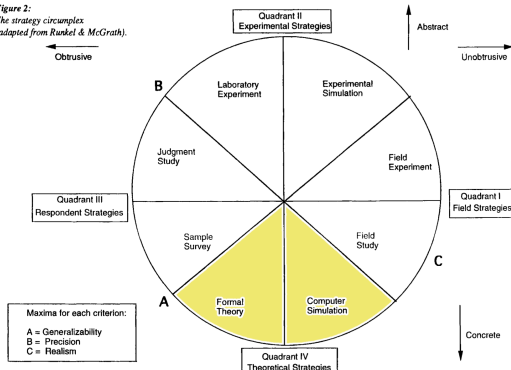
Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



- ▶ Controlled, low realism, high precision, lower generalizability
- ▶ Judgement study: measure the dependent variables concerning the properties of the stimulus

Research Strategies: Everything matters and serves as the building block

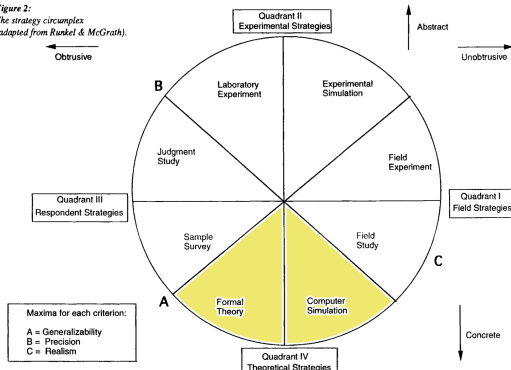
Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



- ▶ Low realism, no precision, high generalizability
- ▶ Formal theory: formulate general rules governing how independent variables affect the dependent variable

Research Strategies: Everything matters and serves as the building block

Figure 2:
The strategy circumplex
(adapted from Runkel & McGrath).



- ▶ High realism, low precision, low generalizability
- ▶ Computer simulation: perform a simulation of the dependent variable in a controllable, self-designed environment

Questions we can ask

Question Type	Example
Base-rate	What is the base-rate of variable Y?
Correlation	Does X correlate with Y? (Remember: correlation $\neq$ causation)
Difference	Does Y take on extreme values when X takes on extreme values?

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give

2. Observations

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ **What are the strengths and weaknesses?**

2. Observations

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Easy to create, cheap, retains most information

2. Observations

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Easy to create, cheap, retains most information
- ▶ Weakness: May change participant behavior

2. Observations

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Easy to create, cheap, retains most information
- ▶ Weakness: May change participant behavior

2. Observations

- ▶ Measure participant behavior of interest

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Easy to create, cheap, retains most information
- ▶ Weakness: May change participant behavior

2. Observations

- ▶ Measure participant behavior of interest
- ▶ What are the strengths and weaknesses?

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Easy to create, cheap, retains most information
- ▶ Weakness: May change participant behavior

2. Observations

- ▶ Measure participant behavior of interest
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Applicable to most situations

Measures we can use

1. Self-reports

- ▶ Summarize answers participants give
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Easy to create, cheap, retains most information
- ▶ Weakness: May change participant behavior

2. Observations

- ▶ Measure participant behavior of interest
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Applicable to most situations
- ▶ Weaknesses: Only overt behavior, costly, may change behavior

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes

4. Trace measures

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ What are the strengths and weaknesses?

4. Trace measures

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Cheap, sometimes the only existing record

4. Trace measures

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ What are the strengths and weaknesses?
- ▶ Strengths: Cheap, sometimes the only existing record
- ▶ Weakness: Can be affected by participant awareness

4. Trace measures

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Cheap, sometimes the only existing record
- ▶ Weakness: Can be affected by participant awareness

4. Trace measures

- ▶ Measure behavior produced but not directly observed by participants

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Cheap, sometimes the only existing record
- ▶ Weakness: Can be affected by participant awareness

4. Trace measures

- ▶ Measure behavior produced but not directly observed by participants
- ▶ **What are the strengths and weaknesses?**

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Cheap, sometimes the only existing record
- ▶ Weakness: Can be affected by participant awareness

4. Trace measures

- ▶ Measure behavior produced but not directly observed by participants
- ▶ **What are the strengths and weaknesses?**
- ▶ Strength: Unobtrusive

Measures we can use (cont.)

3. Analyze historical data

- ▶ Analyze data gathered for other purposes
- ▶ **What are the strengths and weaknesses?**
- ▶ Strengths: Cheap, sometimes the only existing record
- ▶ Weakness: Can be affected by participant awareness

4. Trace measures

- ▶ Measure behavior produced but not directly observed by participants
- ▶ **What are the strengths and weaknesses?**
- ▶ Strength: Unobtrusive
- ▶ Weaknesses: Costly in time, loosely connected to variable

Techniques we can use to manipulate Independent Variables

Selection

Non-random
distribution:
select and assign
cases for each
condition based on
a variable value

Intervention

Random distribu-
tion, chance of
detection: impose
a given value of a
variable through
an overt procedure

Induction

Random distribution,
chance of failure:
induce a given value
of a variable through
a covert procedure

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**
- ▶ External events may affect results

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**
- ▶ External events may affect results

Maturation

- ▶ Participant growth and development
- ▶ Affects longitudinal studies

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**
- ▶ External events may affect results

Selection

- ▶ Non-random assignment

Maturation

- ▶ Participant growth and development
- ▶ Affects longitudinal studies

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**
- ▶ External events may affect results

Selection

- ▶ Non-random assignment
- ▶ **Why could it be a problem?**

Maturation

- ▶ Participant growth and development
- ▶ Affects longitudinal studies

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**
- ▶ External events may affect results

Selection

- ▶ Non-random assignment
- ▶ **Why could it be a problem?**
- ▶ Potential confounding factors

Maturation

- ▶ Participant growth and development
- ▶ Affects longitudinal studies

Threats to Measurements and Variable Manipulation

History

- ▶ Long periods between measurements
- ▶ **Why could it be a problem?**
- ▶ External events may affect results

Maturation

- ▶ Participant growth and development
- ▶ Affects longitudinal studies

Selection

- ▶ Non-random assignment
- ▶ **Why could it be a problem?**
- ▶ Potential confounding factors

Mortality

- ▶ Participant dropout
- ▶ Changes group statistics

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

Selection-Variable Interaction

- ▶ Random selection may interact with other variables
- ▶ E.g., age groups affected by historical events

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

Regression to the mean

- ▶ Extreme values tend to normalize

Selection-Variable Interaction

- ▶ Random selection may interact with other variables
- ▶ E.g., age groups affected by historical events

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

Regression to the mean

- ▶ Extreme values tend to normalize
- ▶ Can be mistaken for intervention effects

Selection-Variable Interaction

- ▶ Random selection may interact with other variables
- ▶ E.g., age groups affected by historical events

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

Regression to the mean

- ▶ Extreme values tend to normalize
- ▶ Can be mistaken for intervention effects
- ▶ **What can we do to omit that?**

Selection-Variable Interaction

- ▶ Random selection may interact with other variables
- ▶ E.g., age groups affected by historical events

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

Selection-Variable Interaction

- ▶ Random selection may interact with other variables
- ▶ E.g., age groups affected by historical events

Regression to the mean

- ▶ Extreme values tend to normalize
- ▶ Can be mistaken for intervention effects
- ▶ **What can we do to omit that?**
- ▶ Assess the random variable:
Run more than one pre-test

More Threats to Measurements and Variable Manipulation

Testing

- ▶ Administer initial testing for, e.g. selection purposes
- ▶ **Why could it be a problem?**
- ▶ Initial testing may affect post-test results

Selection-Variable Interaction

- ▶ Random selection may interact with other variables
- ▶ E.g., age groups affected by historical events

Regression to the mean

- ▶ Extreme values tend to normalize
- ▶ Can be mistaken for intervention effects
- ▶ **What can we do to omit that?**
- ▶ Assess the random variable:
Run more than one pre-test
- ▶ Cross-validate results:
Remove some participants in the analysis

Exercise

Design an experiment to show that our model helps judges make better parole decisions

Failure to Appear: Points		
PSA FACTOR	RESPONSE	POINTS
Pending charge at the time of the arrest	No	0
	Yes	1
Prior conviction (misdemeanor or felony)	No	0
	Yes	1
Prior failure to appear in the past 2 years	No	0
	Yes, just 1	2
	Yes, 2 or more	4
Prior failure to appear older than 2 years	No	0
	Yes	1

- ▶ $x :=$ (prior convictions, severity, age, race, zip code, etc.)
- ▶ Model: $f(x) \rightarrow 1$ iff x should be granted parole

Exercise Questions

- ▶ What is the dependent variable?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?
- ▶ What is the research question?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?
- ▶ What is the research question?
- ▶ Which experimental setup should we use?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?
- ▶ What is the research question?
- ▶ Which experimental setup should we use?
- ▶ What should we definitely control?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?
- ▶ What is the research question?
- ▶ Which experimental setup should we use?
- ▶ What should we definitely control?
- ▶ How would you control for model bias?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?
- ▶ What is the research question?
- ▶ Which experimental setup should we use?
- ▶ What should we definitely control?
- ▶ How would you control for model bias?
- ▶ What confounding variables could we randomize and why?

Exercise Questions

- ▶ What is the dependent variable?
- ▶ How should we measure it?
- ▶ What are the independent variables?
- ▶ What is the research question?
- ▶ Which experimental setup should we use?
- ▶ What should we definitely control?
- ▶ How would you control for model bias?
- ▶ What confounding variables could we randomize and why?
- ▶ What could still go wrong?