On Theory and Identification:

When and Why We Need Theory for Causal Identification

Tara Slough | Columbia/UC Berkeley/NYU

The Identification Revolution and Research Practice

- Increasing **expense** of RHS
  - Field experiments often literally require big $$$
  - Finding “natural experiments” requires effort, expertise

The Identification Revolution and Research Practice

- Increasing **expense** of RHS
  - Field experiments often literally require big $$$
  - Finding “natural experiments” requires effort, expertise

- In many cases, collecting many outcomes is comparatively **cheap**
  - Advocacy for ancillary or downstream experiments
  - Doing a survey is expensive, adding questions is inexpensive

The Identification Revolution and Research Practice

- Increasing **expense** of RHS
  - Field experiments often literally require big $$$
  - Finding "natural experiments" requires effort, expertise
- In many cases, collecting many outcomes is comparatively **cheap**
  - Advocacy for ancillary or downstream experiments
  - Doing a survey is expensive, adding questions is inexpensive
- Implicit **assumption** of this research model:
  - Given some (as-if) random variation in treatment assignment, (some) causal effect is identified for any downstream outcome.
  - This paper shows that this assumption doesn't necessarily hold.
  - Argument: Causal estimands are identified by a design + **theory**.

Argument, in Brief

Z Treatment assignment → **D Treatment** → **Y's Outcomes**

Argument, in Brief

$$\underset{\text{Treatment assignment}}{Z} \longrightarrow \underset{\text{Treatment}}{D} \longrightarrow \underset{\text{Outcomes}}{Y\text{'s}}$$

- Often estimate causal effects on multiple ***Y*'s**; but are not explicit about relationships between ***Y*'s**

Argument, in Brief

$$\begin{tikzcd} \begin{array}{c} Z \\ \textbf{Treatment} \\ \textbf{assignment} \end{array} \arrow[r] & \begin{array}{c} D \\ \textbf{Treatment} \end{array} \arrow[r] & \begin{array}{c} Y\textbf{'s} \\ \textbf{Outcomes} \end{array} \end{tikzcd}$$

- Often estimate causal effects on multiple ***Y*'s**; but are not explicit about relationships between ***Y*'s**

$$\begin{tikzcd}[row sep=small] & & Y_1 \\ Z \arrow[r] & D \arrow[ur] \arrow[dr] & \\ & & Y_2 \end{tikzcd} \quad \text{vs.} \quad \begin{tikzcd} Z \arrow[r] & D \arrow[r] \arrow[rr, bend left] & Y_1 \arrow[r] & Y_2 \end{tikzcd}$$

Argument, in Brief

Z **Treatment assignment** $\longrightarrow$ D **Treatment** $\longrightarrow$ Y**'s Outcomes**

- Often estimate causal effects on multiple Y**'s**; but are not explicit about relationships between Y**'s**

$$\begin{tikzcd} & & Y_1 \\ Z \arrow[r] & D \arrow[ur] \arrow[dr] & \\ & & Y_2 \end{tikzcd} \quad \text{vs.} \quad \begin{tikzcd} Z \arrow[r] & D \arrow[r] \arrow[rr, bend left] & Y_1 \arrow[r] & Y_2 \end{tikzcd}$$

- Relationships between Y**'s** matter for identification and interpretation
- **Argument**: Theory provides assumptions that "discipline" what quantities (estimands) we should estimate, guides interpretation

Definition, Scope of Argument

- **Theory:** A **model**, or an abstract representation of the world, that relies upon deductive reasoning (Clarke and Primo, 2012)
  - Ideally captures: (1) relationships between Z and Y**'s**; (2) relationships between Y**'s**
  - Type of model is not important.
  - Examples here are game- or decision-theoretic.

Definition, Scope of Argument

- **Theory:** A **model**, or an abstract representation of the world, that relies upon deductive reasoning (Clarke and Primo, 2012)
  - Ideally captures: (1) relationships between Z and Y**'s**; (2) relationships between Y**'s**
  - Type of model is not important.
  - Examples here are game- or decision-theoretic.

- **Scope** of argument:
  1. Research designs with (possibly) **sequential** outcomes
     - $\approx$ Dynamic models
  2. Causal estimands as (ideally) **reduced-form** tests of theory
     - Vast majority of common practice

Related Literature

- **Theory vs. identification** debate
  - Huber (2013, 2017); Samii (2016); Ashworth, Berry, and de Mesquita (2015); Clark and Golder (2015)
- Theoretical Implications of Empirical Models (**TIEM**)
  - Eggers (2017); Bueno de Mesquita and Tyson (2019); Sun and Tyson (2019); Prato and Wolton (2019); Izzo, Dewan, and Wolton (2018)
- **Post-treatment** maladies
  - Montgomery, Nyhan, and Torres (2018); Aronow, Baron, and Pinson (2019); Coppock (2019)

Outline

1. **Overview:** Post-treatment selection
2. **Stylized illustration:** Policing experiment
3. **Results:** When we need theory for identification
4. **Implications:** Guidance for research design

Post-Treatment Selection

Truncation by Death

- Consider a clinical trial of treatment Z on a deadly disease
- Ultimate outcome of interest is quality of life, $Y(Z)$
- But some patients die during the trial. Survival denoted $S(Z)$.

Truncation by Death

- Consider a clinical trial of treatment Z on a deadly disease
- Ultimate outcome of interest is quality of life, $Y(Z)$
- But some patients die during the trial. Survival denoted $S(Z)$.

Stratum		$S(1)$	$S(0)$	$\bar{Y}(S=1, Z=1)$	$\bar{Y}(S=1, Z=0)$	$\bar{Y}(S=0, Z=1)$	$\bar{Y}(S=0, Z=0)$
A	π_A	1	1	$\bar{Y}_A(1,1)$	$\bar{Y}_A(1,0)$	-	-
If **T**	π_T	1	0	$\bar{Y}_T(1,1)$	-	-	$\bar{Y}_T(0,0)$
If **U**	π_U	0	1	-	$\bar{Y}_U(1,0)$	$\bar{Y}_U(0,1)$	-
N	π_N	0	0	-	-	$\bar{Y}_N(0,1)$	$\bar{Y}_N(0,0)$

Truncation by Death

- Consider a clinical trial of treatment Z on a deadly disease
- Ultimate outcome of interest is quality of life, $Y(Z)$
- But some patients die during the trial. Survival denoted $S(Z)$.

Stratum		$S(1)$	$S(0)$	$\bar{Y}(S=1, Z=1)$	$\bar{Y}(S=1, Z=0)$	$\bar{Y}(S=0, Z=1)$	$\bar{Y}(S=0, Z=0)$
A	π_A	1	1	$\bar{Y}_A(1,1)$	$\bar{Y}_A(1,0)$	-	-
If **T**	π_T	1	0	$\bar{Y}_T(1,1)$	-	-	Undef.
If **U**	π_U	0	1	-	$\bar{Y}_U(1,0)$	Undef.	-
N	π_N	0	0	-	-	Undef.	Undef.

Truncation by Death

- Consider a clinical trial of treatment Z on a deadly disease
- Ultimate outcome of interest is quality of life, $Y(Z)$
- But some patients die during the trial. Survival denoted $S(Z)$.

Stratum		$S(1)$	$S(0)$	$\bar{Y}(S=1, Z=1)$	$\bar{Y}(S=1, Z=0)$	$\bar{Y}(S=0, Z=1)$	$\bar{Y}(S=0, Z=0)$
A	π_A	1	1	$\bar{Y}_A(1,1)$	$\bar{Y}_A(1,0)$	-	-
If **T**	π_T	1	0	$\bar{Y}_T(1,1)$	-	-	Undef.
If **U**	π_U	0	1	-	$\bar{Y}_U(1,0)$	Undef.	-
N	π_N	0	0	-	-	Undef.	Undef.

$$ATE_S = \pi_A + \pi_T - (\pi_A - \pi_U) = \pi_T - \pi_U \tag{1}$$

$$ATE_Y = \text{undefined} \tag{2}$$

Undefined Potential Outcomes

- Distinct from **missing** data (attrition)
  - Missing data implies that $Y(Z)$ is defined, just not recorded.
  - Methods to address missing data do not "define" a PO.
  - Truncation by death $\rightarrow$ outcomes are **sequential**; not necessary for missing data.

- Usual approach to undefined POs: change the estimand
  - Estimate an (always) survivor average causal effect (**SACE**)
  - Challenge: survivor stratum membership unknowable from data
  - Interval identification via bounds
    - Zhang and Rubin (i.e., 2003)
  - Model-based approaches to point identification; sensitivity analysis etc.

Truncation by Death and Behavioral Outcomes

- In the study of **behavior**:
  - We measure actions taken
  - Past actions can constrain the set of strategies available
  - Common feature of models:
    - e.g., "If x, then game ends, else..."
  - **Sequence** of treatment relative to past actions is critical

Truncation by Death and Behavioral Outcomes

- In the study of **behavior**:
  - We measure actions taken
  - Past actions can constrain the set of strategies available
  - Common feature of models:
    - e.g., "If x, then game ends, else..."
  - **Sequence** of treatment relative to past actions is critical

- Truncation by death = non-availability of a strategy at some **history**
  - Behavior: vote for incumbent (candidate, not party); Selection: Incumbent doesn't run
  - Behavior: behavior after conflict; Selection: Death in conflict

Stylized Illustration: Policing Experiment

Illustration: Policing Experiment

- **Experiment**: See something, say something campaign
  - Treatment: Provision of information about when and how to report
  - Cluster-randomized at the police beat level
  - (Inspired by Colombia Metaketa-IV policing experiment)

Illustration: Policing Experiment

- **Experiment**: See something, say something campaign
  - Treatment: Provision of information about when and how to report
  - Cluster-randomized at the police beat level
  - (Inspired by Colombia Metaketa-IV policing experiment)

- **Mechanism**: Campaign $\rightarrow \downarrow$ net cost of reporting
  - Provides information on how to report
  - Counters social norms <u>against</u> reporting

Illustration: Policing Experiment

- **Experiment**: See something, say something campaign
  - Treatment: Provision of information about when and how to report
  - Cluster-randomized at the police beat level
  - (Inspired by Colombia Metaketa-IV policing experiment)

- **Mechanism**: Campaign $\rightarrow \downarrow$ net cost of reporting
  - Provides information on how to report
  - Counters social norms <u>against</u> reporting

- **Outcomes**: Administrative, behavioral data on:
  - Reporting to 911 equivalent (123): geo-coded, $\approx$ 15k calls/day
  - Crime incidence data: geo-coded, from police records

A Very Simple Model:

- Three **actors**: a suspect, a bystander, and (in paper) a police officer

A Very Simple Model:

- Three **actors**: a suspect, a bystander, and (in paper) a police officer
- Suspect chooses whether or not to commit a crime:
  - Crime yields a benefit $\lambda > 0$, where $\lambda \sim f_\lambda(\cdot)$ with cdf $F_\lambda(\cdot)$
  - If investigated, suspect pays penalty $p > 0$.

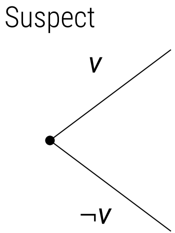
A Very Simple Model:

- Three **actors**: a suspect, a bystander, and (in paper) a police officer
- Suspect chooses whether or not to commit a crime:
  - Crime yields a benefit $\lambda > 0$, where $\lambda \sim f_\lambda(\cdot)$ with cdf $F_\lambda(\cdot)$
  - If investigated, suspect pays penalty $p > 0$.
- <u>If a crime occurs</u>, the bystander reports or does not report
  - If crime is investigated, B receives benefit $\psi > 0$, where $\psi \sim f_\psi(\cdot)$ with cdf $F_\psi(\cdot)$ – a taste for order/justice
  - Reporting incurs cost c_R^Z, where $0 < c_R^{Z=1} < c_R^{Z=0}$

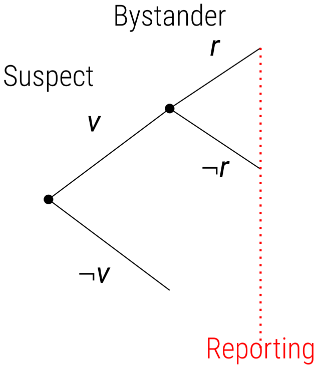
A Very Simple Model:

- Three **actors**: a suspect, a bystander, and (in paper) a police officer
- Suspect chooses whether or not to commit a crime:
  - Crime yields a benefit $\lambda > 0$, where $\lambda \sim f_\lambda(\cdot)$ with cdf $F_\lambda(\cdot)$
  - If investigated, suspect pays penalty $p > 0$.
- <u>If a crime occurs</u>, the bystander reports or does not report
  - If crime is investigated, B receives benefit $\psi > 0$, where $\psi \sim f_\psi(\cdot)$ with cdf $F_\psi(\cdot)$ – a taste for order/justice
  - Reporting incurs cost c_R^Z, where $0 < c_R^{Z=1} < c_R^{Z=0}$
- For presentation, officer is non-strategic
  - w.p. i_R investigates a reported crime and
  - w.p. $i_N < i_R$ investigates an unreported crime

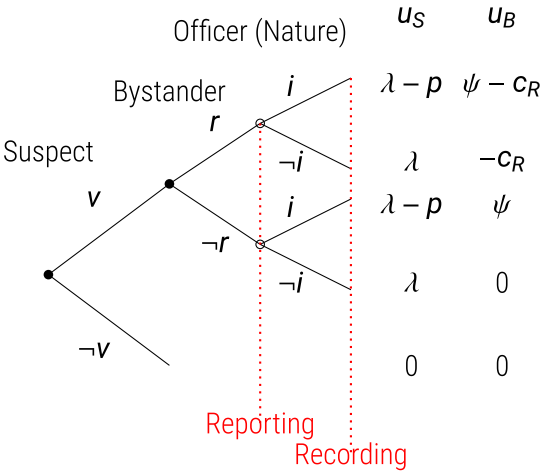
Extensive form



Extensive form



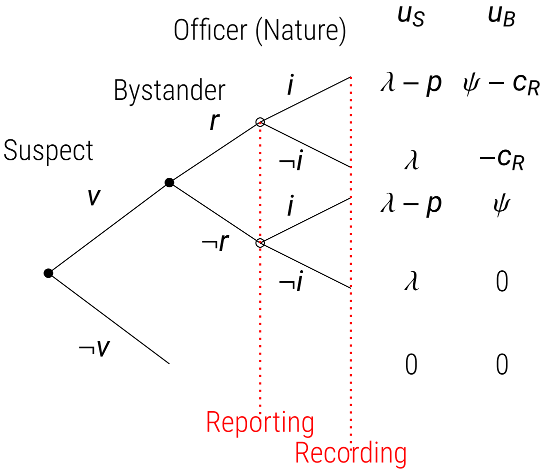
Extensive form



Relevant assumptions:

1. Crime incidence is unmeasured.
2. No false reporting.

Extensive form



- Relevant assumptions:
  1. Crime incidence is unmeasured.
  2. No false reporting

What does the theory teach us?

- Characterize several variants of the model in order to calculate three quantities:
  1. *ATE* of Z on reporting.
  2. *SACE* of Z on reporting.
     - “Survivor” stratum: beats where crime would occur in treatment or control.
  3. Quantity estimated by treatment-control comparison (**difference-in-means** estimator)

Case #1: Always Crime

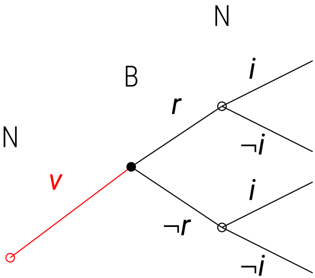


Figure: Always crime.

- Bystander reports if $\psi > \frac{c_R^Z}{\iota_R - \iota_N}$, recall $c_R^{Z=1} < c_R^{Z=0}$
- $ATE = F_\psi\left(\frac{c_R^{Z=0}}{\iota_R - \iota_N}\right) - F_\psi\left(\frac{c_R^{Z=0}}{\iota_R - \iota_N}\right) > 0$
- v always occurs, so $ATE = SACE$.
- DiM estimator estimates ATE.

Case #2: Exogenous Crime

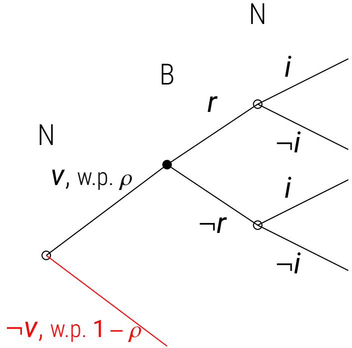


Figure: Exogenous crime. Note that $\rho \in (0, 1)$.

- *ATE* is undefined due to selection into/out of crime
- Like the previous case, $SACE = F_\psi\left(\frac{c_R^{Z=0}}{\iota_R - \iota_N}\right) - F_\psi\left(\frac{c_R^{Z=0}}{\iota_R - \iota_N}\right) > 0$
- …but the DiM estimator estimates $\rho SACE > 0$.

Case #3: Endogenous Crime

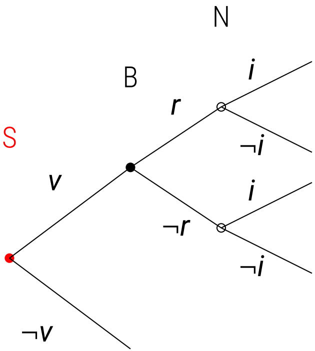


Figure: Endogenous crime.

- *ATE* is undefined due to selection into/out of crime
- Like the previous cases,
  $$SACE = F_\psi\left(\frac{c_R^{Z=0}}{\iota_R - \iota_N}\right) - F_\psi\left(\frac{c_R^{Z=0}}{\iota_R - \iota_N}\right) > 0$$
- DiM estimate incorporates:
  - $\uparrow$ reporting | crime $\rightarrow$ the *SACE*
  - $\downarrow$ crime due to anticipation of reporting
- Ambiguous in sign

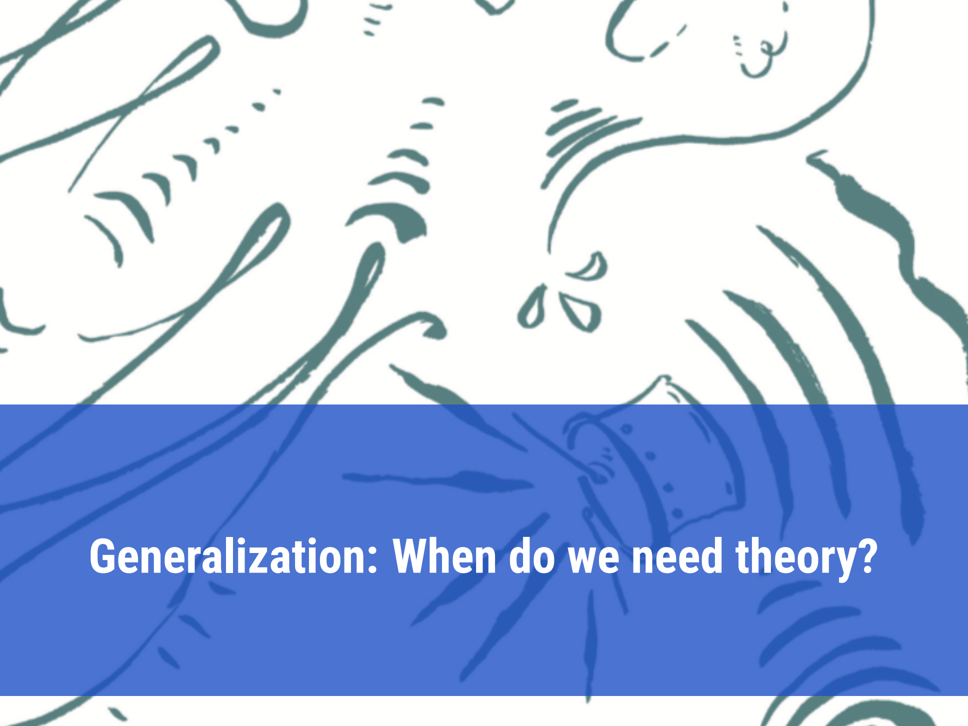
Taking Stock

- Holding research design constant, but changing **theory**:
  - A given estimand (here, the *ATE* on reporting) may or may not be **identified**.
    - Not a “data problem.”
  - **Interpretation** of T vs. C comparison also varies.

Taking Stock

- Holding research design constant, but changing **theory**:
  - A given estimand (here, the *ATE* on reporting) may or may not be **identified**.
    - Not a "data problem."
  - **Interpretation** of T vs. C comparison also varies.

- To what extent are results **general**?
  - Characterize properties of the underlying theory



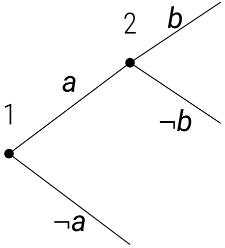
Generalization: When do we need theory?

Strategy Set Symmetry

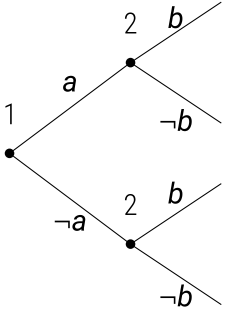
Definition: A model exhibits **strategy set symmetry** if for any history, h, the subsequent actor is the same and has an equivalent strategy set regardless of the strategy selected at h, for all $h \in H \backslash H^T$.

Strategy Set Symmetry

Definition: A model exhibits **strategy set symmetry** if for any history, h, the subsequent actor is the same and has an equivalent strategy set regardless of the strategy selected at h, for all $h \in H \backslash H^T$.



ASYMMETRIC



SYMMETRIC

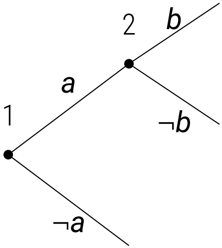
Result: Identification of the *ATE*

Proposition: In an experiment in which standard identifying assumptions hold, if a theory of post-treatment behavior is **not strategy set symmetric**, then:

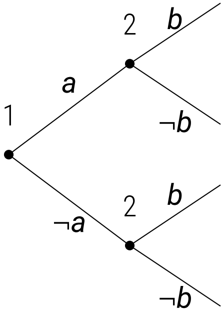
1. There exists at least one post-treatment behavioral outcome for which the *ATE* is identified.
2. There exists at least one post-treatment behavioral outcome for which the *ATE* is not identified.

If a theory of post-treatment behavior is **strategy set symmetric**, then the *ATE* is identified for modeled outcomes.

Graphical Intuition

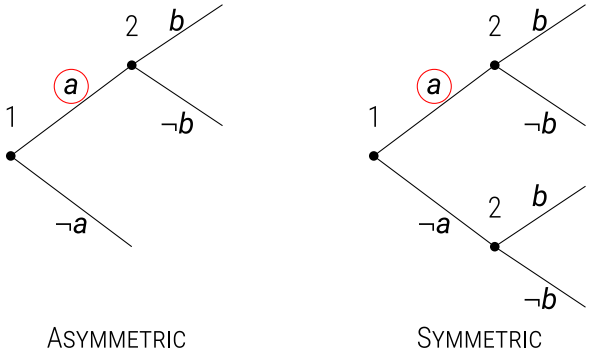


Asymmetric



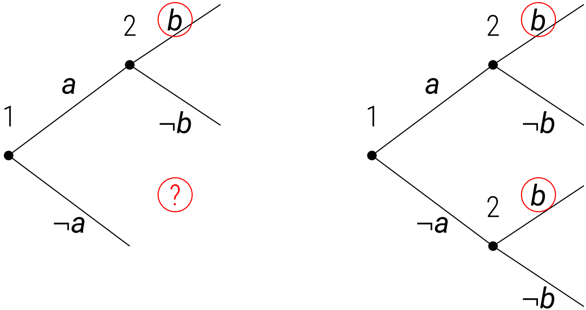
Symmetric

Graphical Intuition



The *ATE* on $A(Z)$ is identified in both cases.

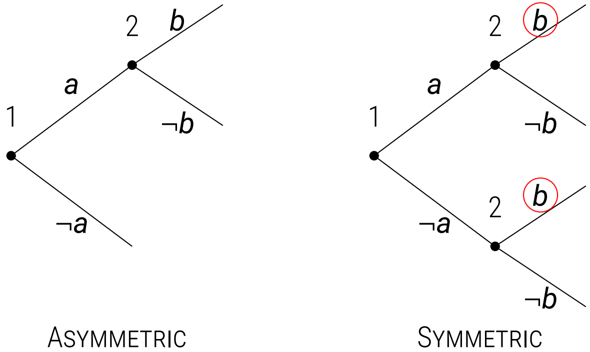
Graphical Intuition



ASYMMETRIC SYMMETRIC

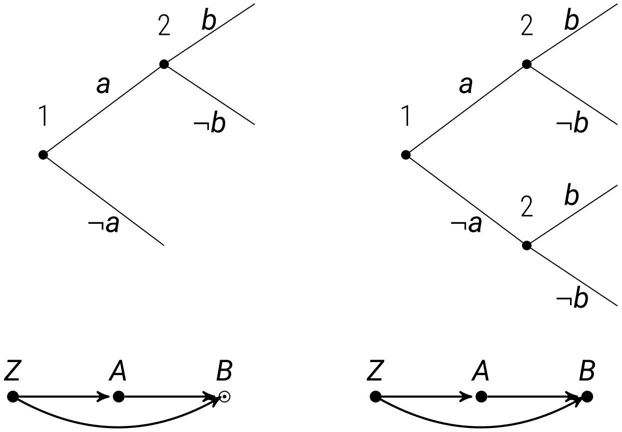
The *ATE* on $B(Z)$ is not identified (left), but is identified (right).

Graphical Intuition



Challenge: the interpretation of the *ATE* on $B(Z)$.

Mapping to a DAG



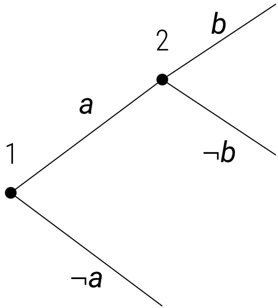
$B(Z)$ is undefined for some levels of $A(Z)$.

Implications

- On theoretical **agnosticism**
  - No theory $\neq$ **agnosticism**
  - Sequential outcomes + claims of identification of (i.e.,) the *ATE*
  - $\Rightarrow$ Strategy set symmetric **model**
- How much of a theory?
  - Identification results hinge on **sequence**, **strategy sets**
  - Nothing about utilities, equilibrium concept etc.
  - Interpretation results greatly assisted by equilibrium/concept of optimal behavior
- On **beliefs** as outcomes
  - Hard to conceive of beliefs as "undefined," even if they are ruled out in equilibrium or via revelation
  - A concern for interpretation, not identification

Recommendations for Research Design

For the Design of Experiments:



1. **Stop** analysis when histories are no longer strategy-set symmetric

2. **Randomize twice** (once per history)
   - No guarantee that partial equilibrium effects tell us anything about the general equilibrium

3. **Flatten** (redefine) outcomes
   - Make Y a **categorical** outcome, $Y \in \{\{a, b\}, \{a, \neg b\}, \neg a\}$
   - Interpretation challenges abound
   - Maybe better with single decision-maker?

Beyond Experiments

- Non-experimental identification-driven research:
  - We observe less of the causal process $\Rightarrow\uparrow$ reliance on assumptions about **extensive form**
  - Longer post-treatment histories $\Rightarrow$ more challenging
  - Recommendations from experiments generally travel but are more demanding: **stop**, **flatten**, or **find** ancillary as-if random variation...

- Other empirical research:
  - Even for **descriptive** purposes, it seems useful to understand where behavioral measures fall in a sequence.

Conclusion: On Theory and Identification

- In the context of **sequential** behavioral outcomes:
  - Identification is a property of a design **+ outcome**
  - Theory guides us to the set of identified estimands
- What if a theory is **wrong?**
  - Can lead to mistaken inferences (or non-inferences)
  - Conception of theory needed for identification is sequence, strategies
  - Absence of a theory $\neq$ no theory
  - Much work to be done here
- Theory vs. identification **debate**:
  - If quest for identification leads us to **underspecify** theory $\Rightarrow$ can be self-defeating.
  - Concern about questions we study.

Thank you!

tara.slough@nyu.edu | www.taraslough.com

Paper available: www.taraslough.com/assets/pdf/theory_id.pdf

Aronow, Peter M., Jonathon Baron, and Lauren Pinson. 2019. "A Note on Dropping Experimental Subjects who Fail a Manipulation Check." *Political Analysis* Forthcoming.

Ashworth, Scott, Christopher R. Berry, and Ethan Bueno de Mesquita. 2015. "All Else Equal in Theory and Data (Big or Small)." *PS Political Science* 48 (1): 89–94.

Bueno de Mesquita, Ethan, and Scott A. Tyson. 2019. "The Commensurability Problem: Conceptual Difficulties in Estimating the Effect of Behavior on Behavior." Working paper.

Clark, William Roberts, and Matt Golder. 2015. "Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?" *PS Political Science* 48 (1): 65–70.

Clarke, Kevin A., and David M. Primo. 2012. *A Model Discipline: Political Science and the Logic of Representations*. New York, NY: Oxford University Press.

Coppock, Alexander. 2019. "Avoiding Post-Treatment Bias in Audit Experiments." *Journal of Experimental Political Science* 6 (1): 1–14.

Eggers, Andrew. 2017. "Quality-Based Explanations of Incumbency Effects." *Journal of Politics* 79 (4): 1315–1328.

Huber, John D. 2013. "Is Theory Getting Lost in the "Identification Revolution"?" *The Political Economist: Newsletter of the Section on Political Economy, American Political Science Association* X (1): 1:3.

Huber, John D. 2017. *Exclusion by Elections: Inequality, Ethnic Identity, and Democracy*. New York: Cambridge University Press.

Izzo, Federica, Torun Dewan, and Stephane Wolton. 2018. "Cumulative Knowledge in the Social Sciences: The Case of Improving Voters' Information." Working Paper available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3239047`.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–775.

Prato, Carlo, and Stephane Wolton. 2019. "Electoral Imbalances and their Consequences." *Journal of Politics* First View: 1–15.

Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *Journal of Politics* 78 (3): 941–955.

Sun, Jessica S., and Scott A. Tyson. 2019. "Theoretical Implications of Empirical Models: An Application to Conflict Studies." Working paper.

Zhang, Junni L., and Donald B. Rubin. 2003. "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by "Death"." *Journal of Educational and Behavioral Statistics* 28 (4): 353–368.