This talk is based on the following paper:


Help from Dustin Tingley is acknowledged
Causal inference is a central goal of social science
Randomized experiments as **gold standard**
But, experiments are a **black box**
Can only tell *whether* the treatment causally affects the outcome
Not *how* and *why* the treatment affects the outcome

Challenge is how to identify **causal mechanisms**

---

**What This Talk is About**

- **Goal**: Show how to identify causal mechanisms using statistics
- **Method**: Causal Mediation Analysis

![Diagram](attachment:image.png)

- Direct and indirect effects; intermediate and intervening variables
- Popular among social psychologists (e.g., Baron and Kenny)
Current Practice

- Regression
  \[ Y_i = \alpha + \beta T_i + \gamma M_i + \delta X_i + \epsilon_i \]
- Each coefficient is interpreted as a causal effect
- Sometimes, it’s called marginal (or partial) effect
- Idea: increase \( T_i \) by one unit while holding \( M_i \) and \( X_i \) constant

- Post-treatment bias: if you change \( T_i \), that may also change \( M_i \)
- Usual advice: only include causally prior (or pre-treatment) variables
- But, then you lose causal mechanisms!

Causal Mediation in Interactive Experiment I

- Communication can influence behavior in strategic games
- But what psychological mechanisms are at work?
- Drolet and Morris (2000). *J. of Experimental Social Psychology*
- Rapport vs. positive affect and expectations

![Diagram of Causal Mediation](image_url)
Experimental Design and Finding

- How does rapport differ from positive affect and expectations?
  - shared sense of mutual understanding
  - dyadic or group level process
  - convergence of nonverbal expressions
  - observable by a third party

- Experimental Design:
  - randomized treatment: face-to-face or telephone conversation
  - talk about “positive experiences at Stanford”
  - fill out surveys measuring rapport and positive expectations
  - measure outside observers’ perception of rapport
  - play a single shot PD game

- Finding: rapport mediates the positive effects of face-to-face communication, but positive affect and expectations do not

Causal Mediation Analysis in Interactive Experiment II

- People overbid in auctions. Why?
- Useful for designing better auctions
- Delgado et al. (2008) Science
- Fear of losing vs. Joy of winning

[Diagram showing causal mediation analysis with arrows indicating treatment, mediator, and outcome variables.]
Experimental Design and Findings

• Randomized treatment: lottery or two-person auction
• fMRI to measure BOLD response to outcomes in the right striatum
• Evaluate causal mechanisms of overbidding

• Greater change in BOLD signal when subject lost in auction
• Little change when subject won
• Important mediating role of fear of losing in auction

Formal Statistical Framework of Causal Inference

• Units: $i = 1, \ldots, n$
• “Treatment”: $T_i = 1$ if treated, $T_i = 0$ otherwise
• Observed outcome: $Y_i$
• Pre-treatment covariates: $X_i$
• Potential outcomes: $Y_i(1)$ and $Y_i(0)$ where $Y_i = Y_i(T_i)$

<table>
<thead>
<tr>
<th>Subject pair $i$</th>
<th>Communication type $T_i$</th>
<th>Cooperation $Y_i(1)$</th>
<th>$Y_i(0)$</th>
<th>Average age $X_{1i}$</th>
<th>How many economists $X_{2i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>21.5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>?</td>
<td>1</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>22</td>
<td>2</td>
</tr>
</tbody>
</table>

• Causal effect: $Y_i(1) - Y_i(0)$
Identification of Causal Effects in Standard Settings

- **Nonparametric identification**: What can we learn from the data generating process alone?
- **Average Treatment Effect (ATE)**:
  \[ \tau \equiv \mathbb{E}(Y_i(1) - Y_i(0)) \]
- **Ignorability (randomization, no omitted variable)**:
  \[ (Y_i(1), Y_i(0)) \perp \perp T_i \mid X_i \]
- **Identification under ignorability**:
  \[ \tau = \mathbb{E}(Y_i \mid T_i = 1, X_i) - \mathbb{E}(Y_i \mid T_i = 0, X_i) \]
- **Relationship with the linear regression**:
  \[ Y_i(T_i) = \alpha + \beta T_i + \gamma X_i + \epsilon_i \]
  where ignorability implies \( T_i \perp \perp \epsilon_i \mid X_i \)

Notation for Causal Mediation Analysis

- **Binary treatment**: \( T_i \in \{0, 1\} \)
- **Mediator**: \( M_i \)
- **Outcome**: \( Y_i \)
- **Observed covariates**: \( X_i \)
- **Potential mediators**: \( M_i(t) \) where \( M_i = M_i(T_i) \)
- **Potential outcomes**: \( Y_i(t, m) \) where \( Y_i = Y_i(T_i, M_i(T_i)) \)
Defining and Interpreting Causal Mediation Effects

- **Total causal effect:** \( \tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \)

- **Causal mediation effects:**
  \( \delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \)

- Change the mediator from \( M_i(0) \) to \( M_i(1) \) while holding the treatment constant at \( t \)

- Indirect effect of the treatment on the outcome through the mediator under treatment status \( t \)

- \( Y_i(t, M_i(t)) \) is observable but \( Y_i(t, M_i(1 - t)) \) is not

- **Direct effects:**
  \( \zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \)

- Change the treatment from 0 to 1 while holding the mediator constant at \( M_i(t) \)

- Total effect = mediation (indirect) effect + direct effect:
  \( \tau_i = \delta_i(t) + \zeta_i(1 - t) \)

- **Quantities of interest:** Average Causal Mediation Effects,
  \( \bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t)) = \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\} \)
The Main Identification Result

Assumption 1 (Sequential Ignorability)

\[ \{ Y_i(t, m), M_i(t) \} \perp T_i \mid X_i, \]
\[ Y_i(t, m) \perp M_i \mid T_i, X_i \]

for \( t = 0, 1 \) and \( m \in \mathcal{M} \).

Theorem 1 (Nonparametric Identification)

Under Assumption 1, for \( t = 0, 1 \),

\[ \tilde{\delta}(t) = (-1)^t \int \{ \int \mathbb{E}(Y_i \mid M_i, T_i = t, X_i) \, dP(M_i \mid T_i = 1 - t, X_i) \]
\[ - \mathbb{E}(Y_i \mid T_i = t, X_i) \} \, dP(X_i) \]

Theoretical and Practical Implications

- Existing statistics literature concludes that an additional assumption is required for the identification of mediation effects.
- However, sequential ignorability alone is sufficient.
- Fit two nonparametric regressions:
  1. \( \mu_{tm}(x) \equiv \mathbb{E}(Y_i \mid T_i = t, M_i = m, X_i = x) \)
  2. \( \lambda_{tm}(x) \equiv \text{Pr}(M_i = m \mid T_i = t, X_i = x) \)
- The plug-in estimator for a discrete mediator:

\[
(-1)^t \left\{ \sum_{m=0}^{J-1} \sum_{i=1}^{n} 1\{T_i = 1 - t\} \hat{\lambda}_{1-t,m}(X_i) \sum_{i=1}^{n} 1\{T_i = t\} \hat{\mu}_{tm}(X_i) \hat{\lambda}_{tm}(X_i) \\
\frac{1}{n_{1-t}} \sum_{i=1}^{n} 1\{T_i = t\} \hat{\lambda}_{tm}(X_i) \\
- \frac{1}{n_t} \sum_{i=1}^{n} 1\{T_i = t\} \left( \sum_{m=0}^{J-1} \hat{\mu}_{tm}(X_i) \hat{\lambda}_{tm}(X_i) \right) \right\}.
\]
Identification under Linear Structural Equation Model

Theorem 2 (Identification under LSEM)

Consider the following linear structural equation model

\[
M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \\
Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}.
\]

Under Assumption 1, the average causal mediation effects are identified as \( \bar{\delta}(0) = \bar{\delta}(1) = \beta_2 \gamma \).

Run two (not three) regressions and multiply two coefficients!

- Direct effect: \( \beta_3 \)
- Total effect: \( \beta_2 \gamma + \beta_3 \)
- Total effect could be zero even when mediation effects are not

Need for Sensitivity Analysis

- The sequential ignorability assumption is often too strong
- Need to assess the robustness of findings via sensitivity analysis
- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?
- Parametric and nonparametric sensitivity analysis by assuming

\[
\{ Y_i(t, m), M_i(t) \} \perp \perp T_i \mid X_i
\]

but not

\[
Y_i(t, m) \perp \perp M_i \mid T_i, X_i
\]
- **Sensitivity parameter**: \( \rho \equiv \text{Corr}(\epsilon_{2i}, \epsilon_{3i}) \)
- Existence of omitted variables leads to non-zero \( \rho \)
- Set \( \rho \) to different values and see how mediation effects change
- All you have to do: fit another regression

\[
Y_i = \alpha_3^* + \beta_3^* T_i + \epsilon_{3i}^*
\]

in addition to the previous two regressions:

\[
M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}
\]

\[
Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}
\]

- Estimated causal mediation effects as a function of \( \rho \) (and identifiable parameters)

**Theorem 3 (Identification with a Given Error Correlation)**

*Under Assumption 3,*

\[
\bar{\delta}(0) = \bar{\delta}(1) = \beta_2 \left( \frac{\sigma_{23}^*}{\sigma_2^2} - \frac{\rho}{\sigma_2} \sqrt{\frac{1}{1 - \rho^2} \left( \frac{\sigma_{3i}^*}{\sigma_3^2} - \frac{\sigma_{23}^*}{\sigma_2^2} \right)} \right),
\]

where \( \sigma_j^2 \equiv \text{Var}(\epsilon_{ji}) \) for \( j = 2, 3, \sigma_{3i}^* \equiv \text{Var}(\epsilon_{3i}^*), \sigma_{23}^* \equiv \text{Cov}(\epsilon_{2i}, \epsilon_{3i}^*), \) and \( \epsilon_{3i}^* = \gamma \epsilon_{2i} + \epsilon_{3i}. \)

- When do my results go away completely?
- \( \bar{\delta}(t) = 0 \) if and only if \( \rho = \text{Corr}(\epsilon_{2i}, \epsilon_{3i}^*) \) (easy to compute!)
How does media framing affect citizens’ political opinions?
- News stories about the Ku Klux Klan rally in Ohio
- Free speech frame ($T_i = 0$) and public order frame ($T_i = 1$)
- Randomized experiment with the sample size = 136

- Mediators: general attitudes (12 point scale) about the importance of free speech and public order
- Outcome: tolerance (7 point scale) for the Klan rally
- Expected findings: negative mediation effects
### Analysis under Sequential Ignorability

<table>
<thead>
<tr>
<th>Mediator</th>
<th>Public Order</th>
<th>Free Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimator</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parametric</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-interaction</td>
<td>$-0.510$</td>
<td>$-0.126$</td>
</tr>
<tr>
<td>$\hat{\delta}(0)$</td>
<td>$-0.451$</td>
<td>$-0.131$</td>
</tr>
<tr>
<td>$\hat{\delta}(1)$</td>
<td>$-0.566$</td>
<td>$-0.122$</td>
</tr>
<tr>
<td><strong>Nonparametric</strong></td>
<td>$-0.374$</td>
<td>$-0.094$</td>
</tr>
<tr>
<td>$\hat{\delta}(0)$</td>
<td>$-0.374$</td>
<td>$-0.094$</td>
</tr>
<tr>
<td>$\hat{\delta}(1)$</td>
<td>$-0.596$</td>
<td>$-0.222$</td>
</tr>
</tbody>
</table>

### Parametric Sensitivity Analysis

#### Parametric Analysis

![Diagram showing sensitivity analysis](image-url)
Concluding Remarks

- Quantitative analysis can be used to identify causal mechanisms!
- Estimate causal mediation effects rather than marginal effects
- Wide applications in social science disciplines