Text as Causal Mediators:
Research Design for Causal Estimates of Differential Treatment of Social Groups via Language Aspects

Katherine A. Keith, Douglas Rice, and Brendan O'Connor

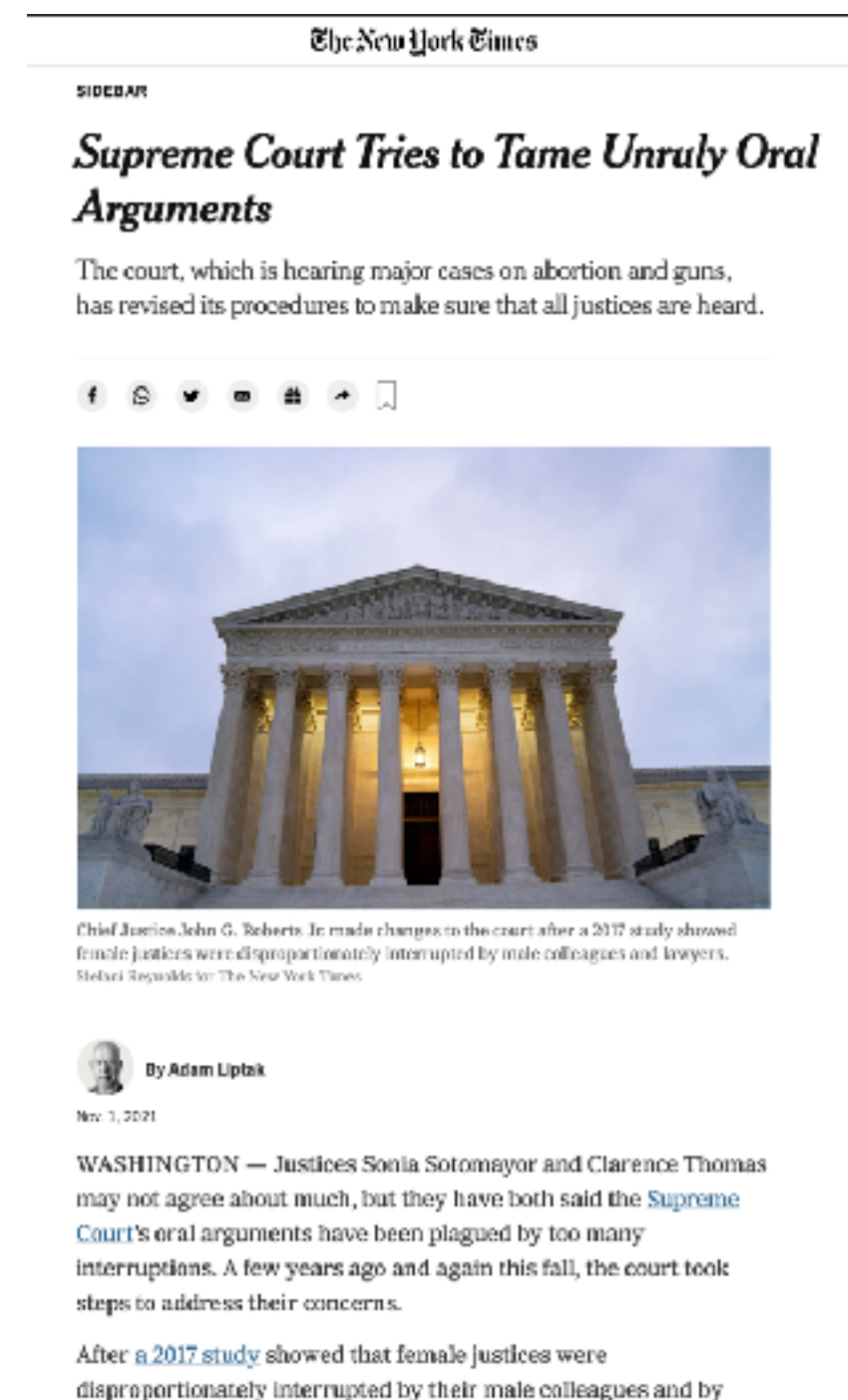# Bias in interruptions during U.S. Supreme Court oral arguments

Q: Why do some justices interrupt female advocates more than male advocates?

(Patton & Smith, "Lawyer, Interrupted: Gender Bias in Oral Arguments at the U.S. Supreme Court," *Journal of Law and Courts*, 2017)

(Jacobi and Schweers. "Justice, interrupted: The effect of gender, ideology, and seniority at Supreme Court oral arguments." Va. L. Rev, 2017)

Importance of interruptions as causal outcome

- Interruptions => status reinforcement (Mendelberg et al., 2014)

- Justices' oral argument behavior <=> case outcomes (Johnson et al., 2006)

- Timely and relevant

Bias in interruptions during U.S. Supreme Court oral arguments

Q: Why do some justices interrupt female advocates more than male advocates?

Legal analysts

Explanation 1: Implicit gender bias

- Different types of clients with weaker legal arguments
- Decreased quality of the argument
- Manner of speaking

Explanation 2: Women are "less effective" advocates

(Patton & Smith, "Lawyer, Interrupted: Gender Bias in Oral Arguments at the U.S. Supreme Court," *Journal of Law and Courts*, 2017)

(Jacobi and Schweers. "Justice, interrupted: The effect of gender, ideology, and seniority at Supreme Court oral arguments." Va. L. Rev, 2017)

Example

Lozano v. Montoya Alvarez (2013)

Ann O'Connell Adams (advocate):

Well—

Antonin Scalia (justice):

I mean, it seems to me it just makes that article impossible to apply consistently country to country.

Ann O'Connell Adams (advocate):

No, I don't think so. And—and, the other signatories have—have almost all, I mean I think the Hong Kong court does say that it doesn't have discretion, but […] the other courts of signatory countries that have interpreted Article 12 have all found a discretion, whether it be in Article 12 or in Article 8.—

Antonin Scalia (justice):

Have they exercised it? Have they exercised it, that discretion which they say is there?

Example
Lozano v. Montoya Alvarez (2013)

Ann O'Connell Adams (advocate):
(Photo Credit: LinkedIn)

Well— **Interruption**

Antonin Scalia (justice):
(Photo Credit: Brookings Institute)

I mean, it seems to me it just makes that article impossible to apply consistently country to country.

Hedging **Speech Disfluencies**

Ann O'Connell Adams (advocate):

No, I don't think so. And—and, the other signatories have—have almost all, I mean I think the Hong Kong court does say that it doesn't have discretion, but […] the other courts of signatory countries that have interpreted Article 12 have all found a discretion, whether it be in Article 12 or in Article 8.— **Interruption**
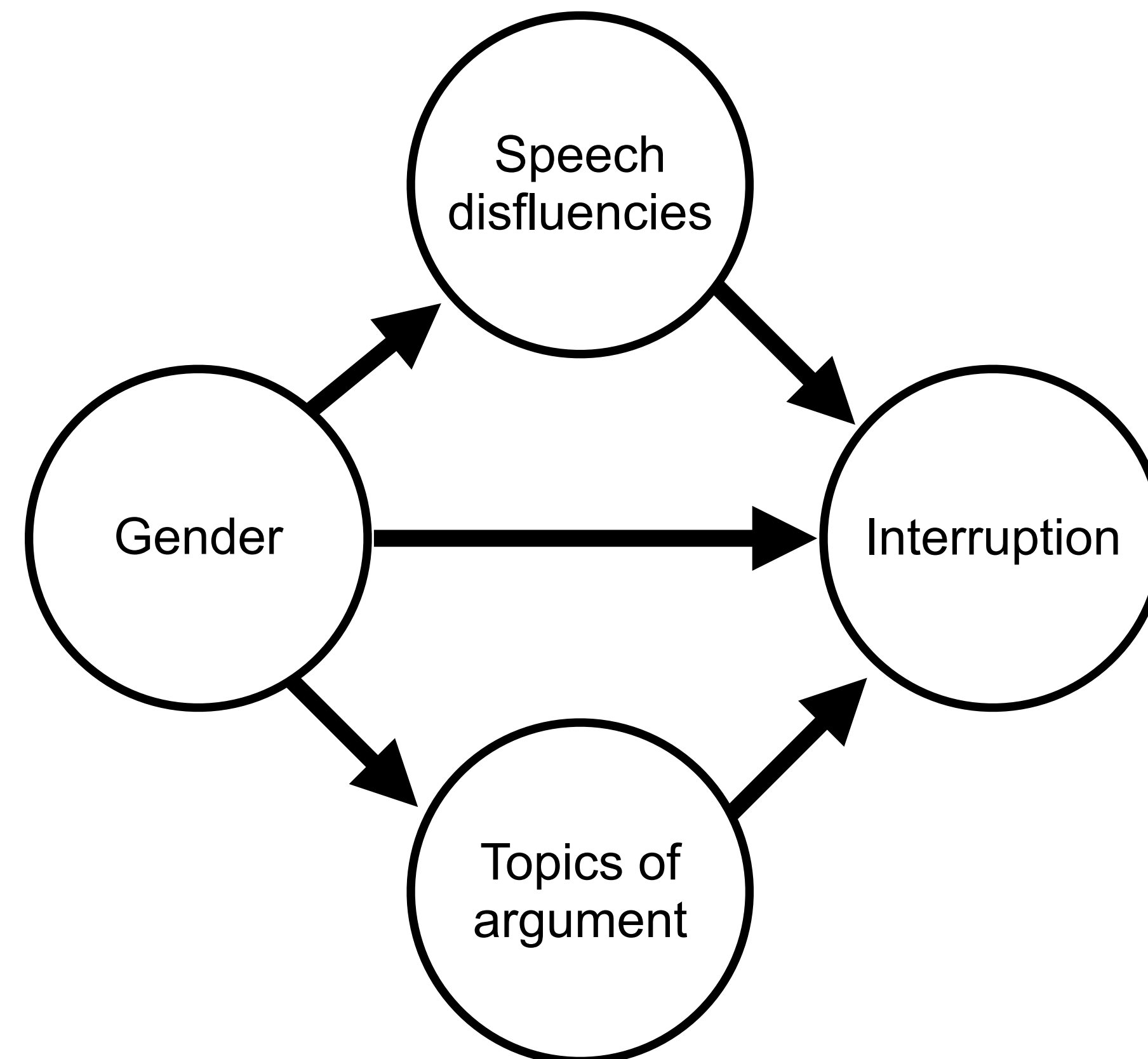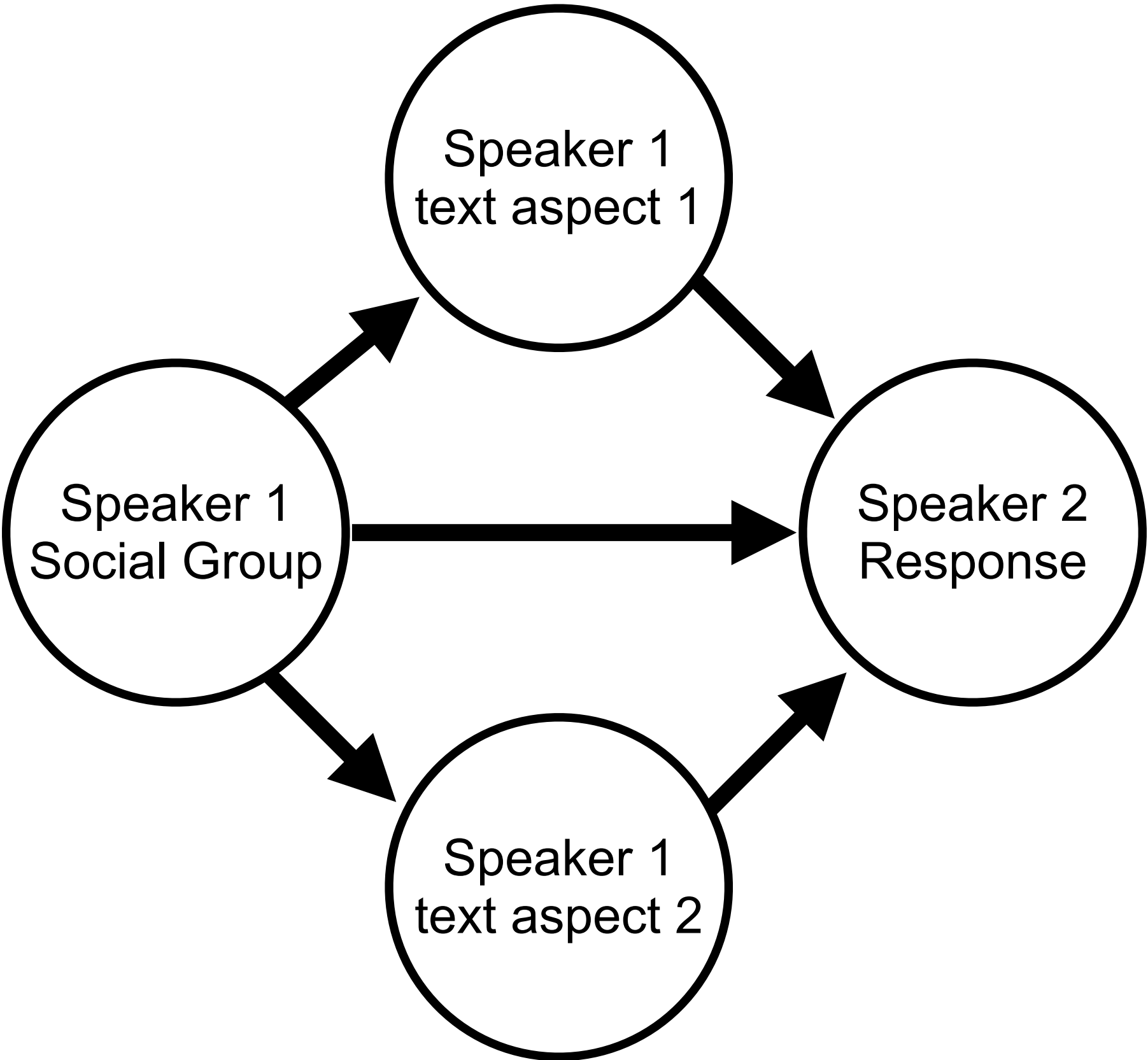
Antonin Scalia (justice):

Have they exercised it? Have they exercised it, that discretion which they say is there?
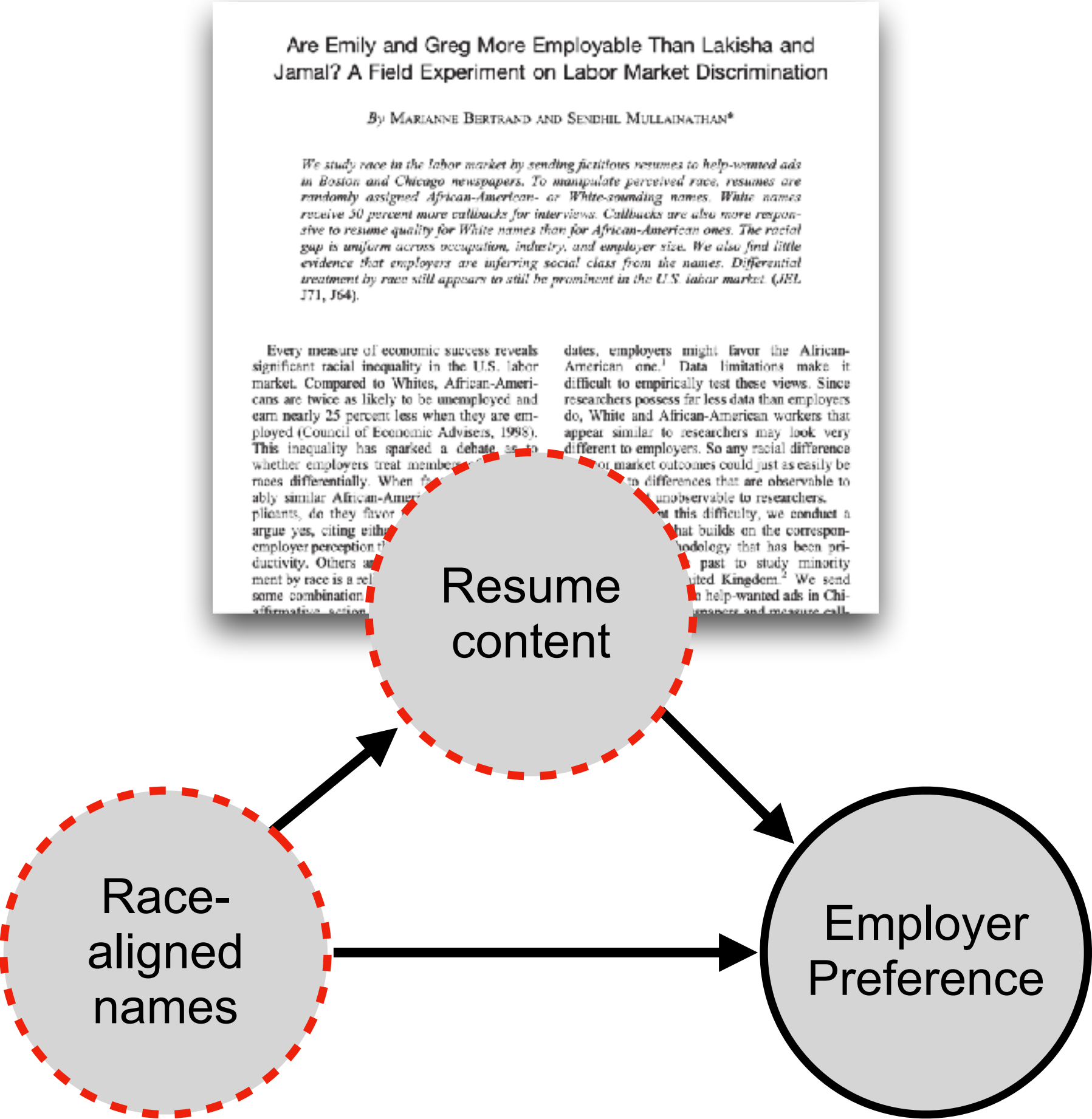
Causal DAG, U.S. Supreme Court

Causal DAG, General Framework

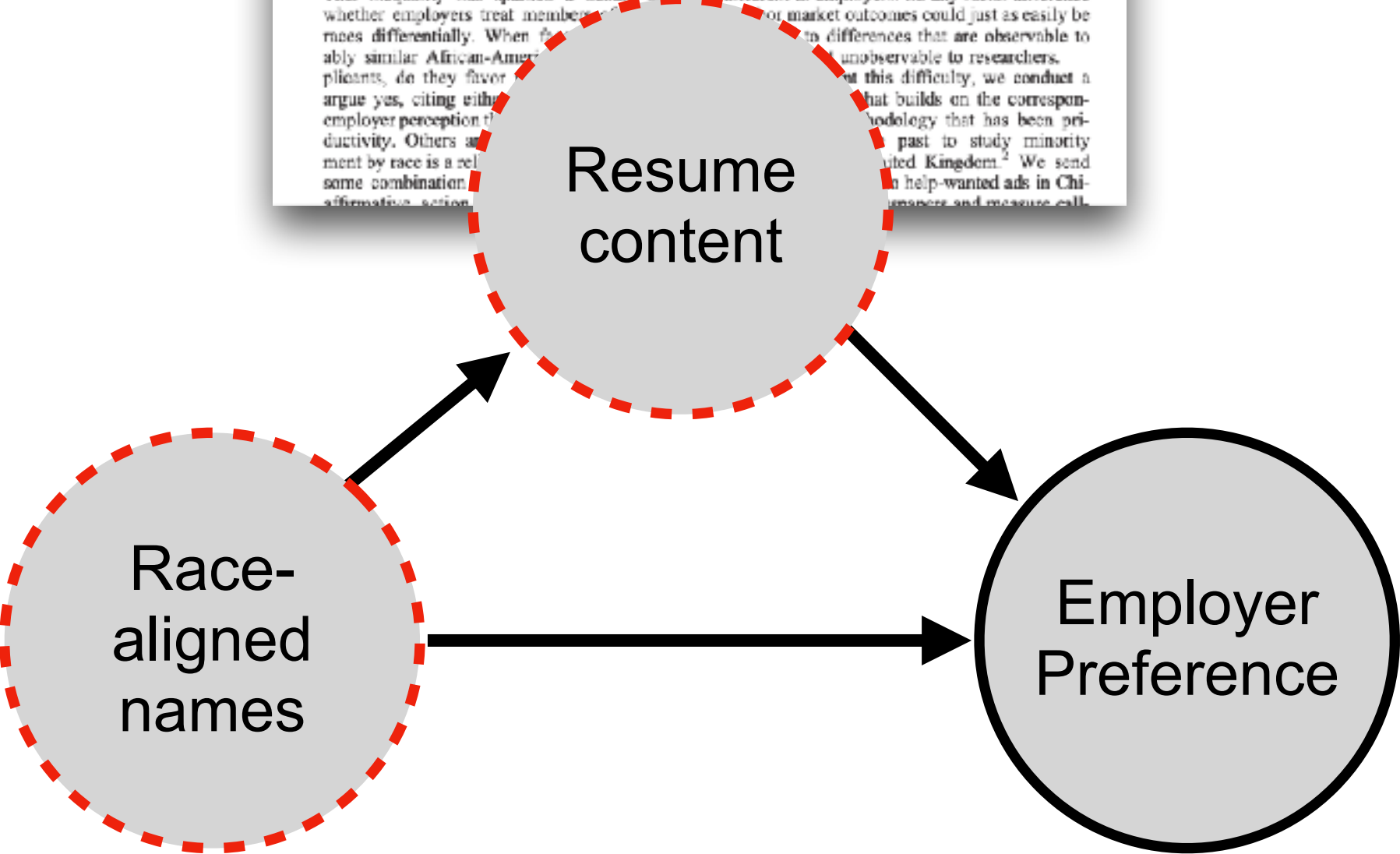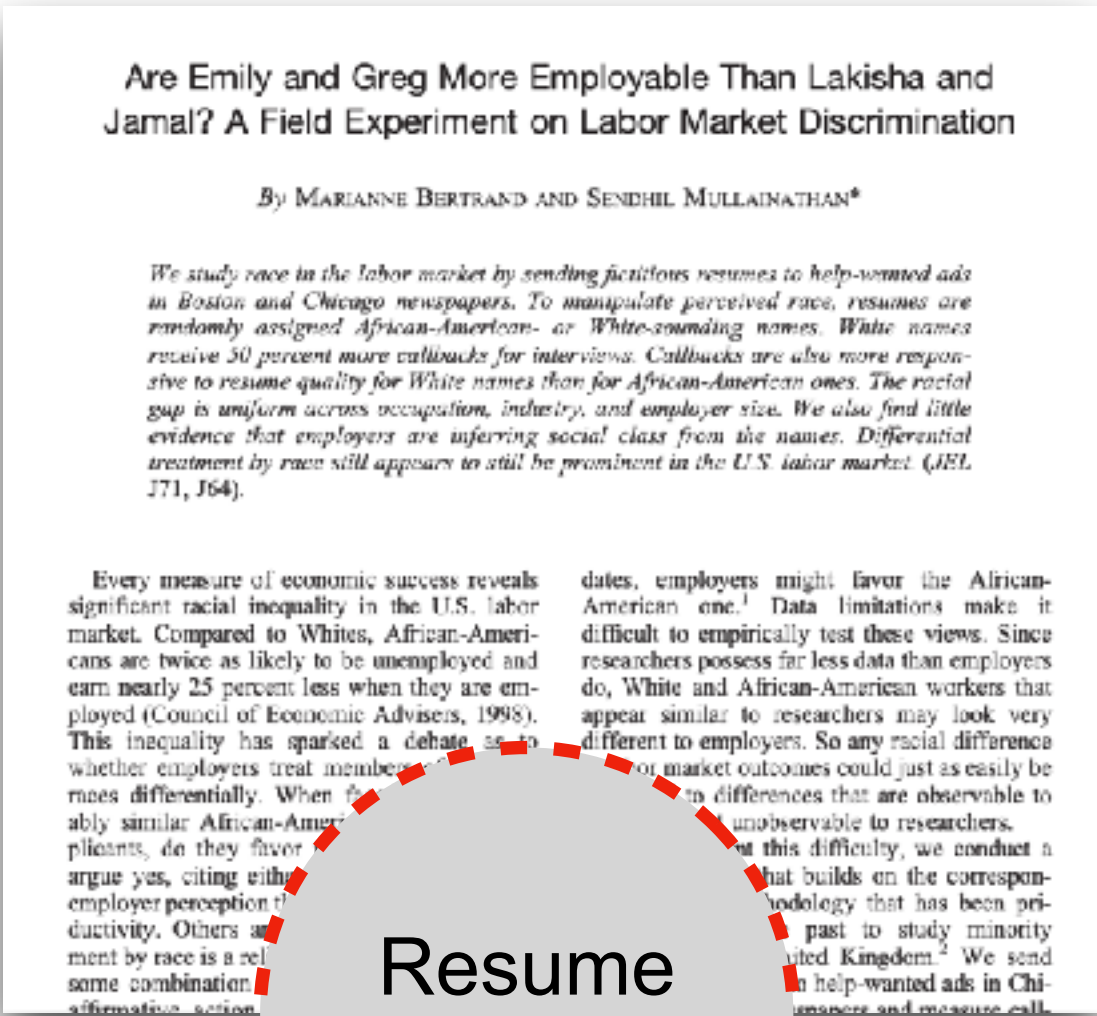# Causal experiments (audit studies) with bias + text

Causal experiments (audit studies) with bias + text

Resumes
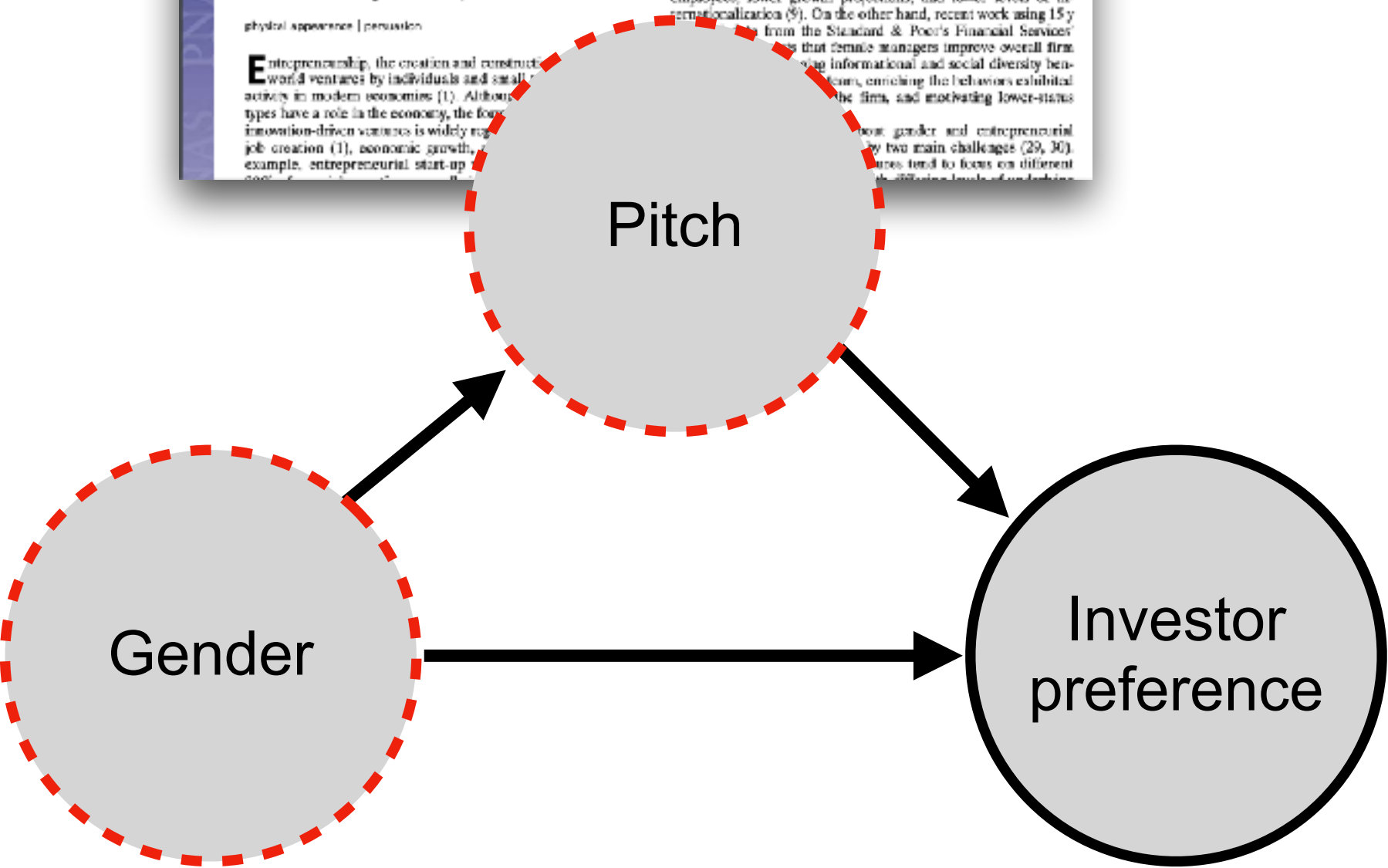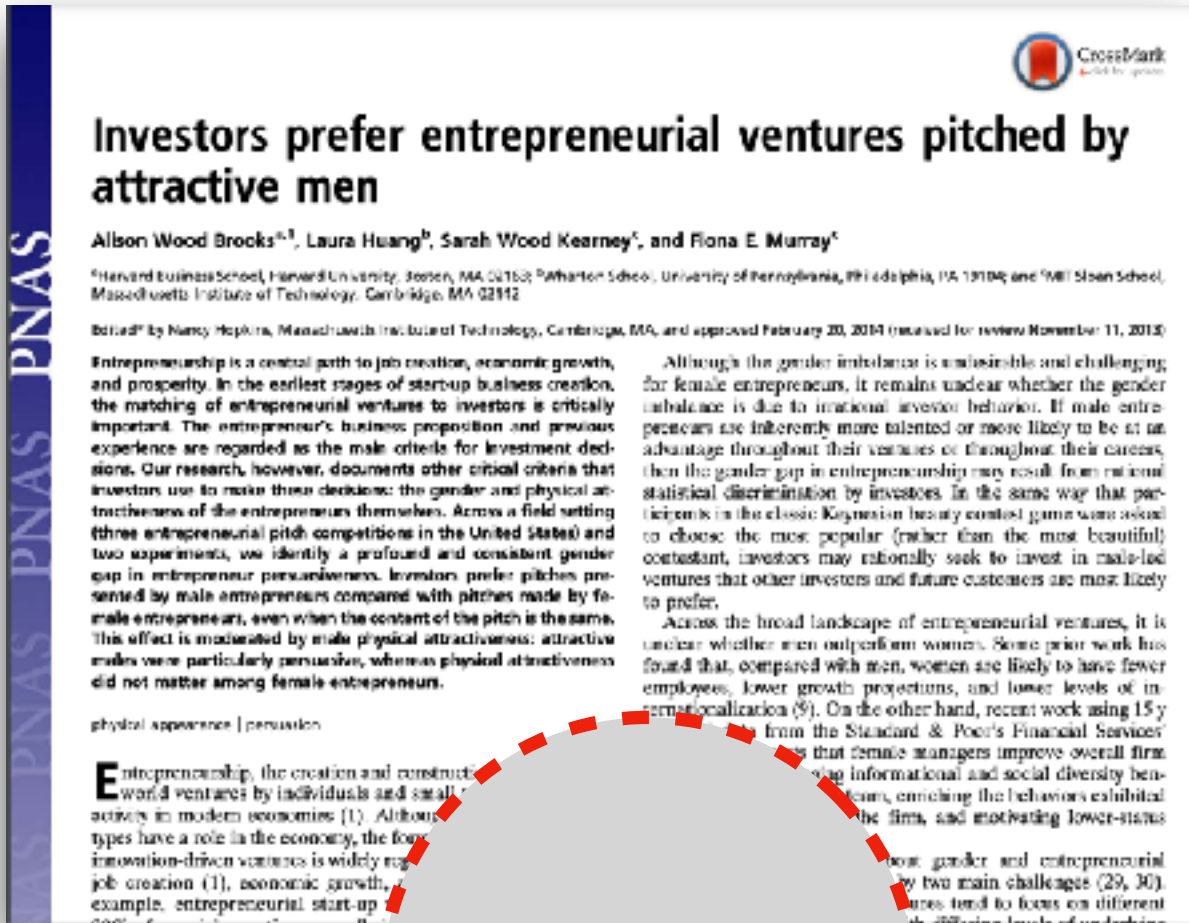
Causal experiments (audit studies) with bias + text

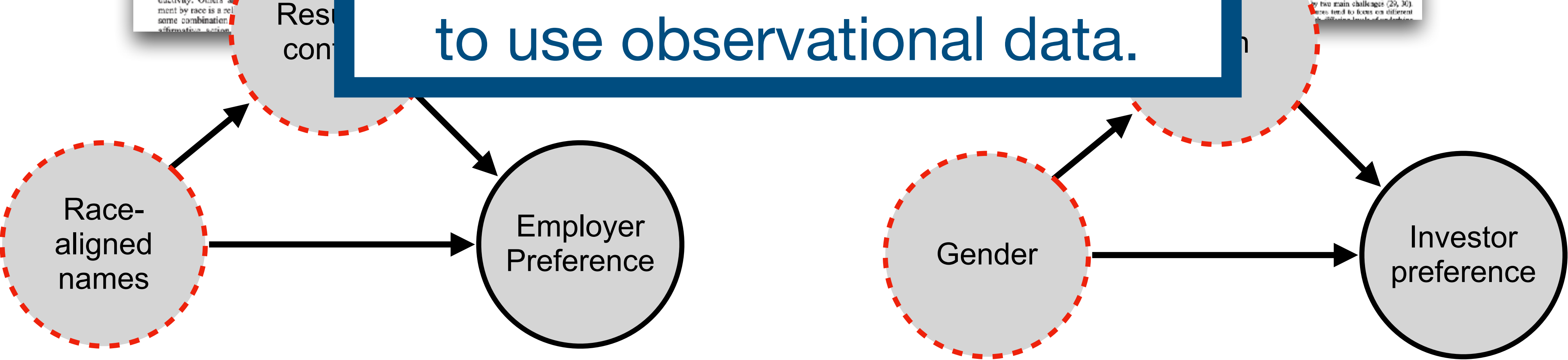## Resumes

Entrepreneurial pitches

Causal experiments (audit studies) with bias + text
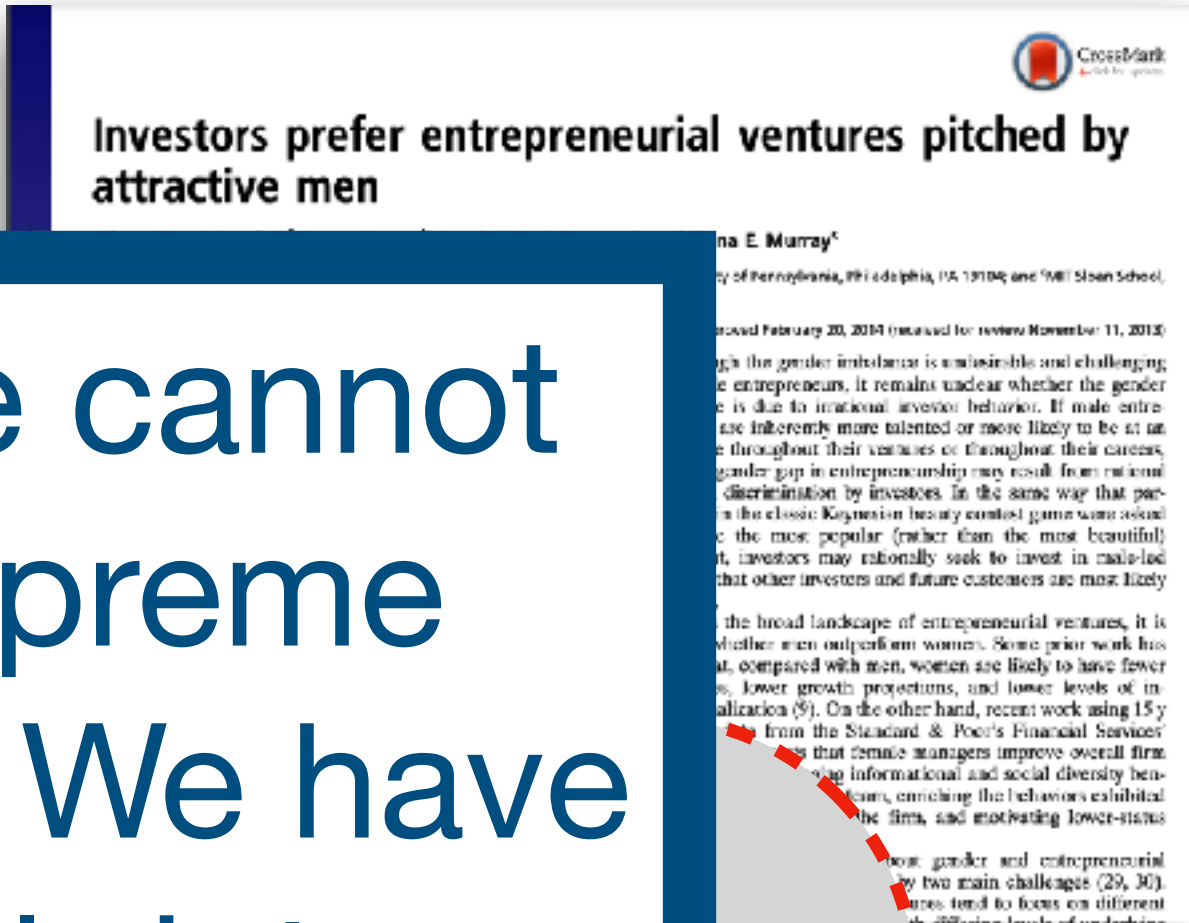
Resumes

Entrepreneurial pitches

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

By MARIANNE BERTRAND AND SENDHIL MULLAINATHAN*

We study race in the labor market by sending fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perception of race, each resume is randomly assigned either a very African-American sounding name or a very White sounding name. The results show that White names receive 50 percent more callbacks for interviews. Callbacks are also more responsive to resume quality for White names than for African-American ones. The racial gap is uniform across occupation, industry, and employer size. We also find little evidence that employers are inferring social class from the names. Differential treatment by race still appears to still be prominent in the U. S. labor market. (JEL J15, J64).

Every measure of economic success reveals significant racial inequality in the U.S. labor market. Compared to Whites, African-Americans are twice as likely to be unemployed and earn nearly 25 percent less when they are employed (Council of Economic Advisers, 1998). This inequality has sparked a debate as to whether employers treat members of different races differently. When faced with observably similar African-American and White applicants, do they favor the White one? Some argue yes, citing either employer prejudice or employer perception that race signals lower productivity. Others argue that differential treatment by race is a relic of the past, eliminated by some combination of employer enlightenment, affirmative action ...

Investors prefer entrepreneurial ventures pitched by attractive men

Major challenge: We cannot *intervene* during Supreme Court oral arguments. We have to use observational data.

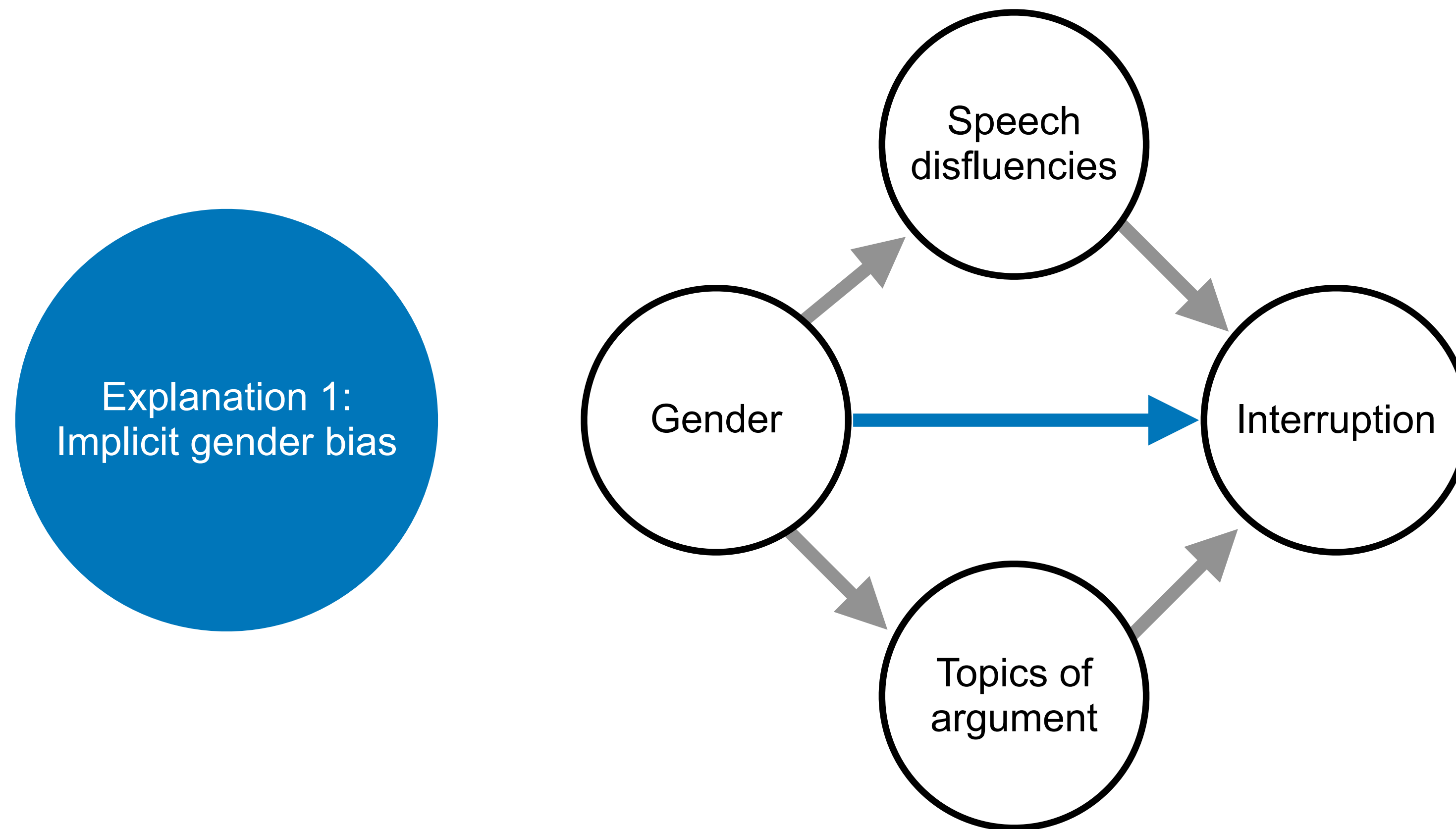Race-aligned names → Employer Preference

Gender → Investor preference

Contributions & past work context

- Intentionally focusing on a thoughtful **causal design** before we obtain empirical results

  - "Design trumps analysis" (Rubin, 2008)

  - We will only every have observational data for the U.S. Supreme Court

- We use **causal mediation analysis** towards the goal of splitting the total effect into the portion of the effect that goes through language mediators and the portion that does not

  - General causal mediation analysis: (Pearl, 2001; Imai et al., 2010; VanderWeele, 2016)

  - Text and mediation at this workshop: (Tierney & Volfovsky, 2021)

- Illustrate the **challenges** conceptualizing and operationalizing causal variables

  - Criticisms of claiming "gender" or "race" as a causal treatments (Sen & Wasow, 2016; Hu & Kohler-Hausmann, 2020)

  - Difficult to choose which language aspects to choose as mediators (e.g. Pryzant et al., 2021 with text as treatment)
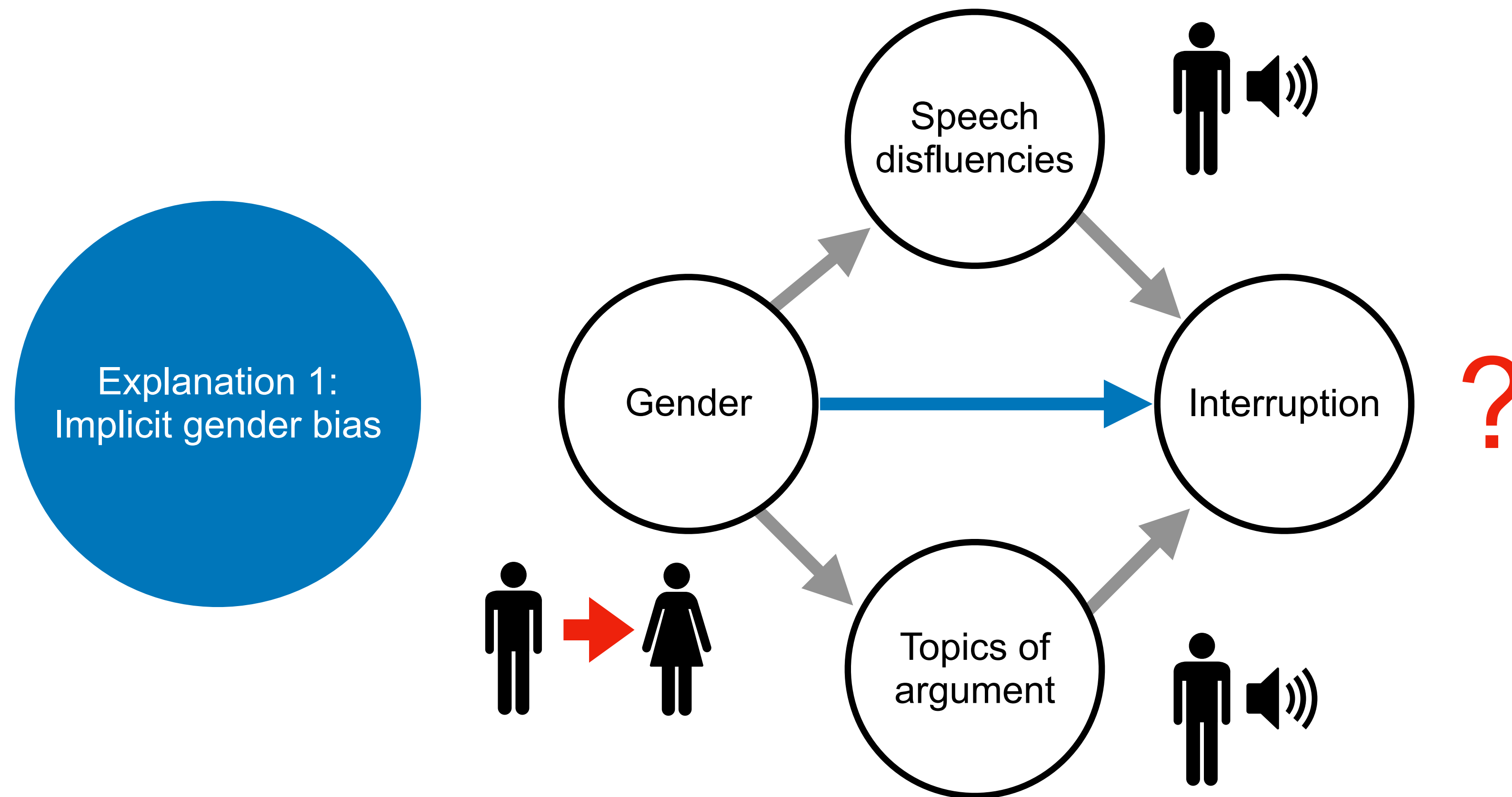
Causal DAG, U.S. Supreme Court
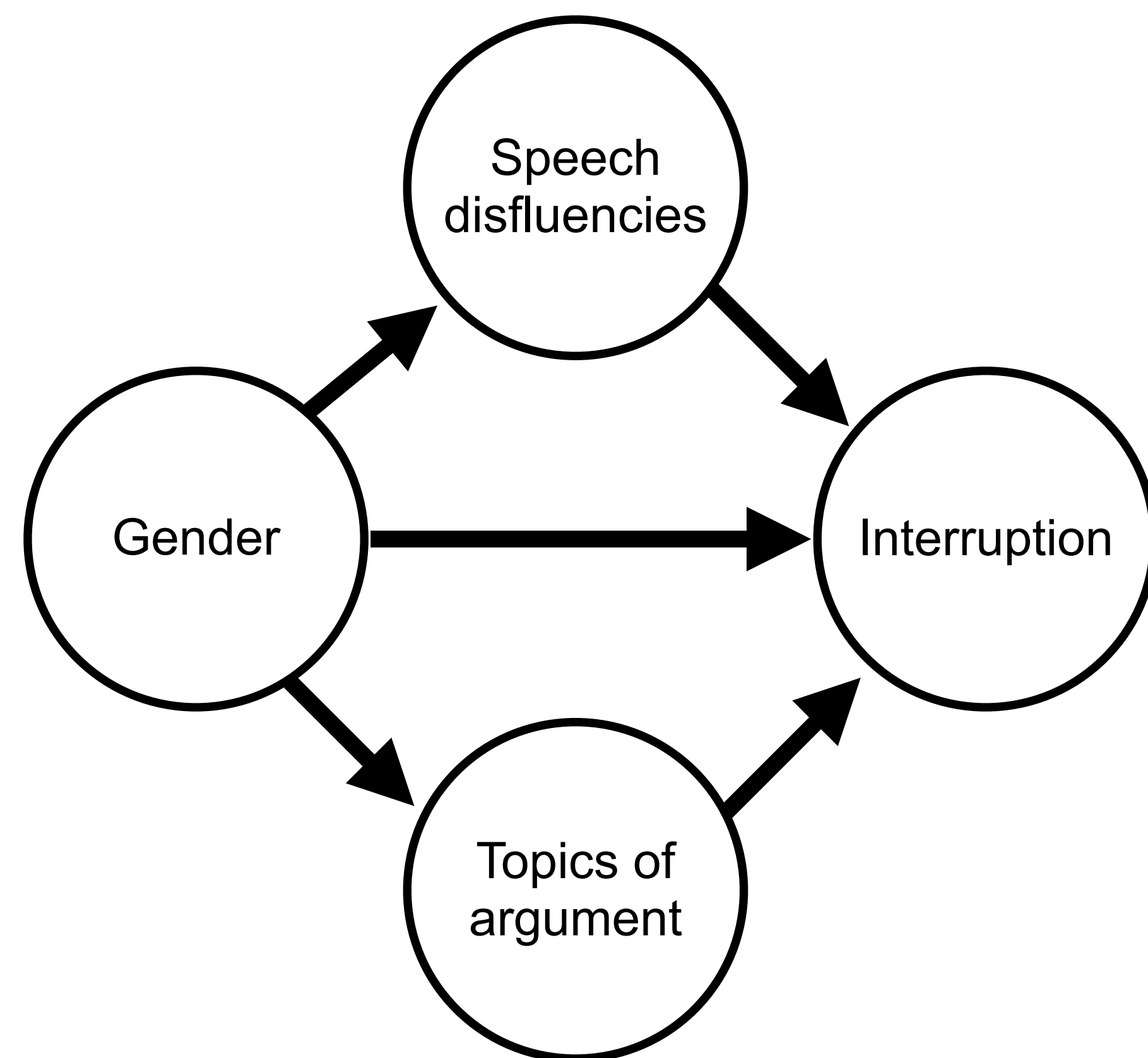
Explanation 1 corresponds to the direct path

Explanation 1 corresponds to the direct path

Explanation 1: Implicit gender bias

Speech disfluencies

Gender

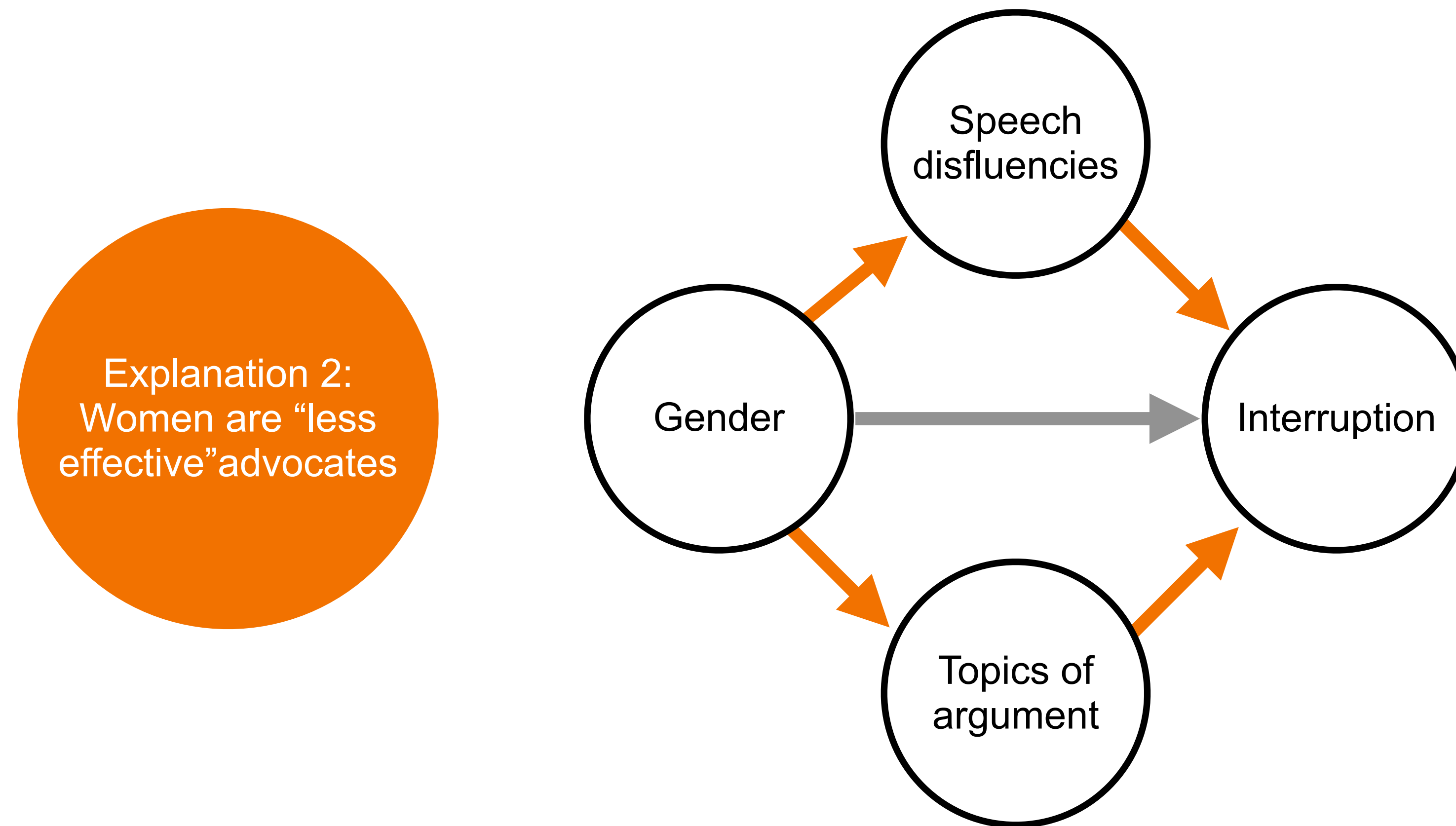Topics of argument

Interruption

?

Natural direct effect (NDE)

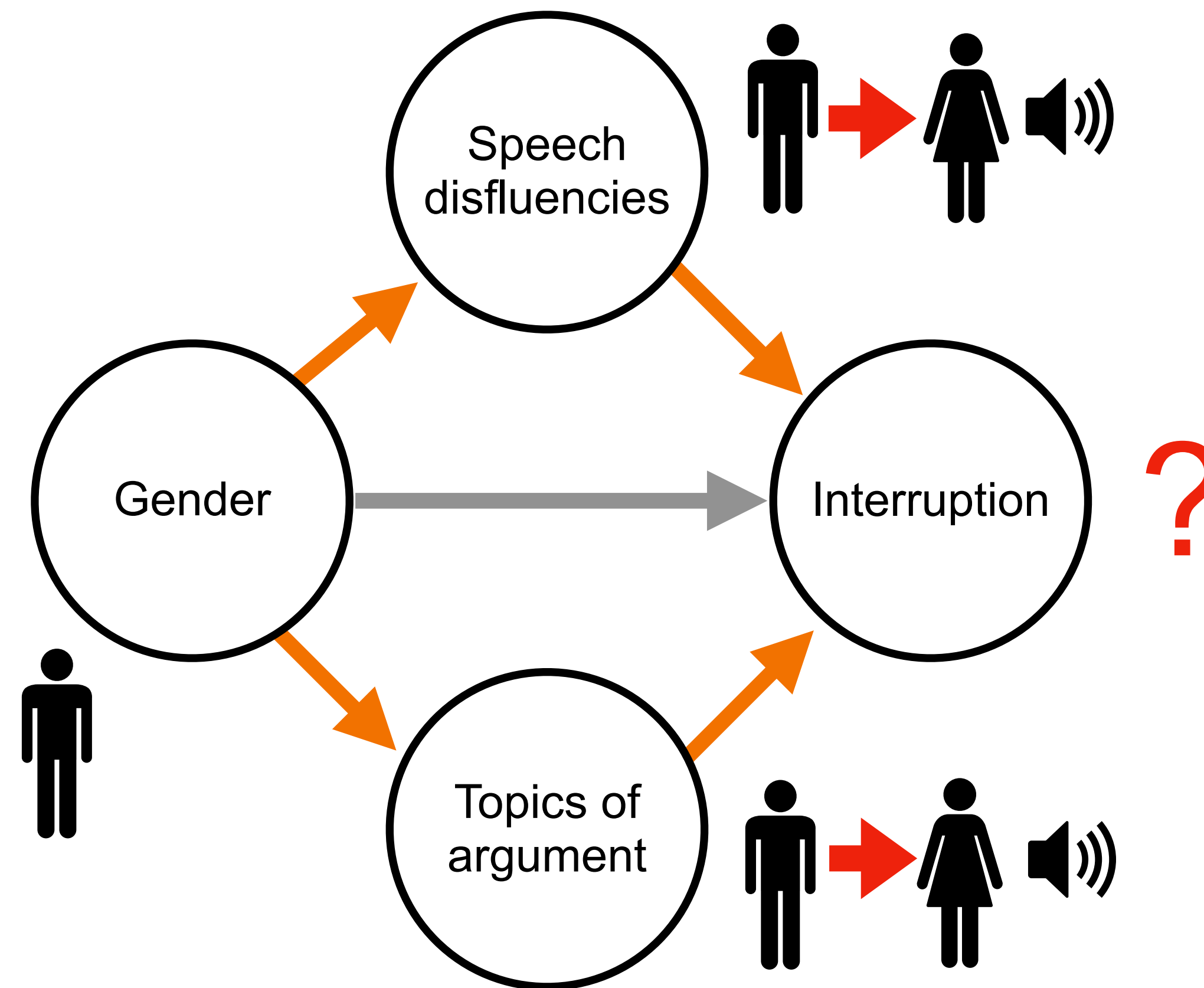How would a justice's interruptions of an advocate change if
- the signal of the advocate's gender the justice received flipped from male to female
- but the advocate still used language typical of a male advocate?

Explanation 2 corresponds to paths through mediators

Explanation 2 corresponds to paths through mediators

Explanation 2: Women are "less effective" advocates

Natural indirect effect (NIE)

How would a justice's interruptions of an advocate change if

- a male advocate used language typical of a female advocate
- but the signal of the advocate's gender the justice received remained male?

Identification and Estimation

Based on Imai et al. 2010 and Pearl et al. 2016

- Sequential ignorability
  (Imai et al, 2010)

$$\{Y_i(t', m), M_i\}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) \mid \{T_i = t, X_i = x\}$$

- Mediator Independence

$$\forall j, j' : \; M_i^j(t) \perp\!\!\!\perp M_i^{j'}(t) \mid \{T_i = t, X_i = x\}$$
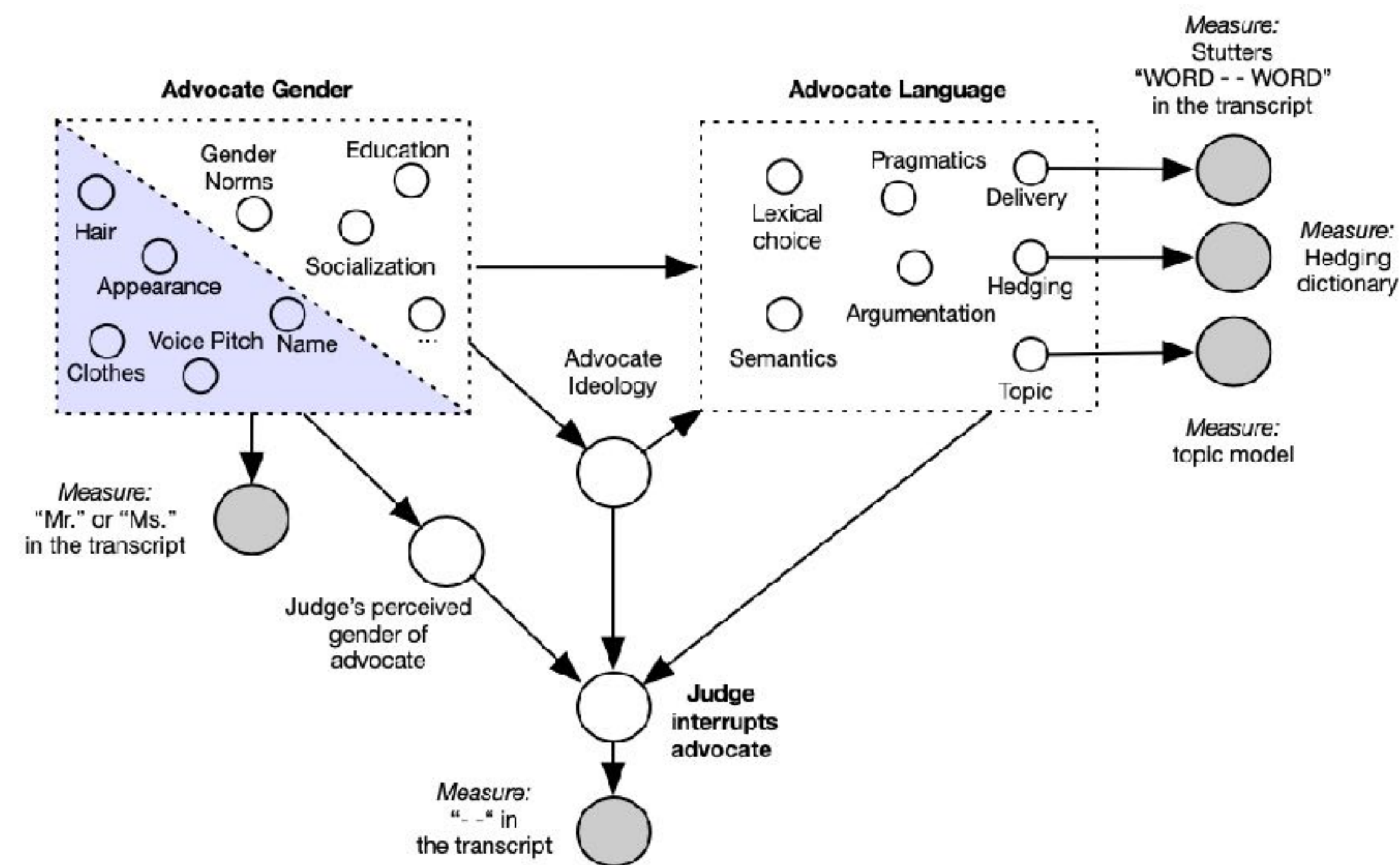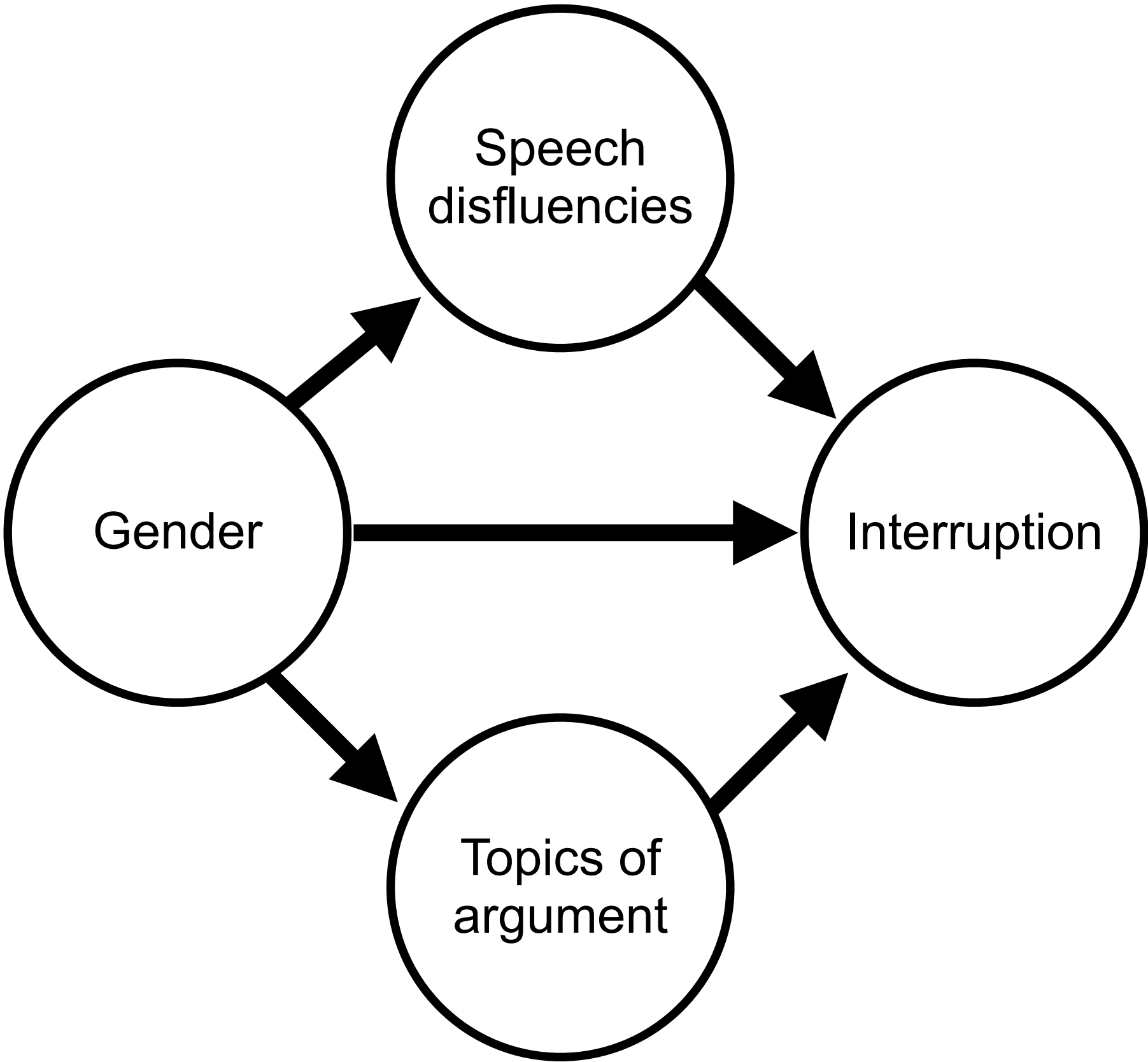
$$\text{SA-NDE}^j =$$

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{x \in \mathcal{X}} \sum_{m \in \mathcal{M}^j} \left(\hat{f}^j(Y | M_i^j = m, T_i = 1, X_i = x) \right.$$

$$\left. , X_i = x) \right) \hat{g}^j(m | T_i = 0, X_i = x)$$

$$\text{SA-NIE}^j =$$

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{x \in \mathcal{X}} \sum_{m \in \mathcal{M}^j} \hat{f}^j(Y | M_i^j = m, T_i = 0, X_i = x)$$

$$\left(\hat{g}^j(m | T_i = 1, X_i = x) - \hat{g}^j(m | T_i = 0, X_i = x) \right)$$
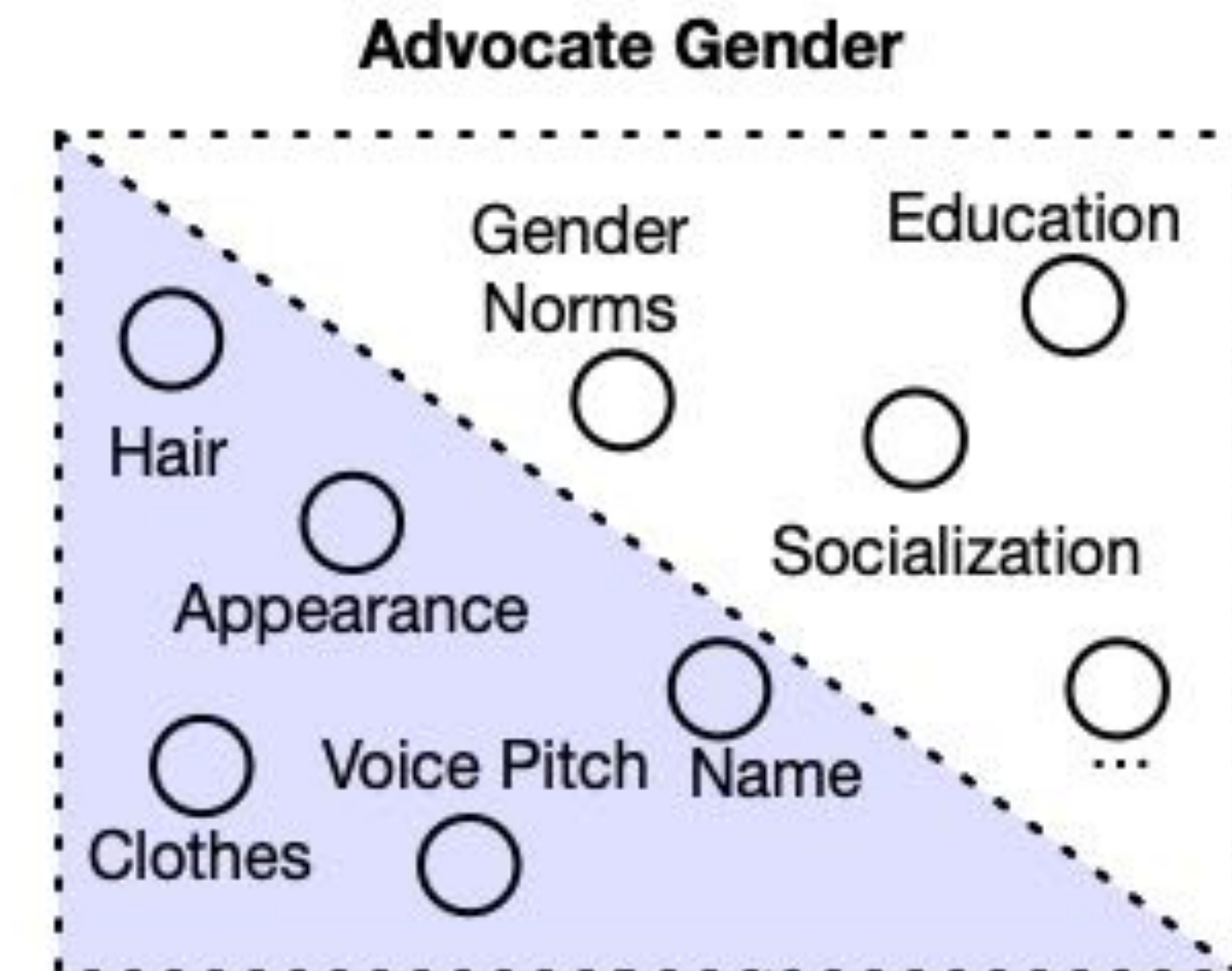
See paper and poster

Limitations of simple assumptions

Gender as a causal "treatment"

Treatment options

1. Do judges interrupt at different rates based on an advocate's *gender*?

2. Based on an advocate's *biological sex assigned at birth*?

3. An advocate's *perceived gender*?

4. An advocate's *gender signal*?

5. **An advocate's *gender signal* as defined by (hypothetical) manipulations of the advocate's clothes, hair, name, and voice pitch?**

6. An advocate's *gender signal* by (hypothetical) manipulations of their entire physical appearance, facial features, name, and voice pitch?

7. An advocate's *gender signal* by setting their physical appearance, facial features, name, and voice pitch to specific values (e.g. all facial features set to that of the same 40-year-old, white female and clothes set to a black blazer and pants).

Advocate Gender

Hair · Appearance · Gender Norms · Education · Socialization · Voice Pitch · Name · Clothes · ...

Building from Sen and Wasow (2016); Hu and Kohler-Hausmann (2020)

Operationalizing language as a causal mediator
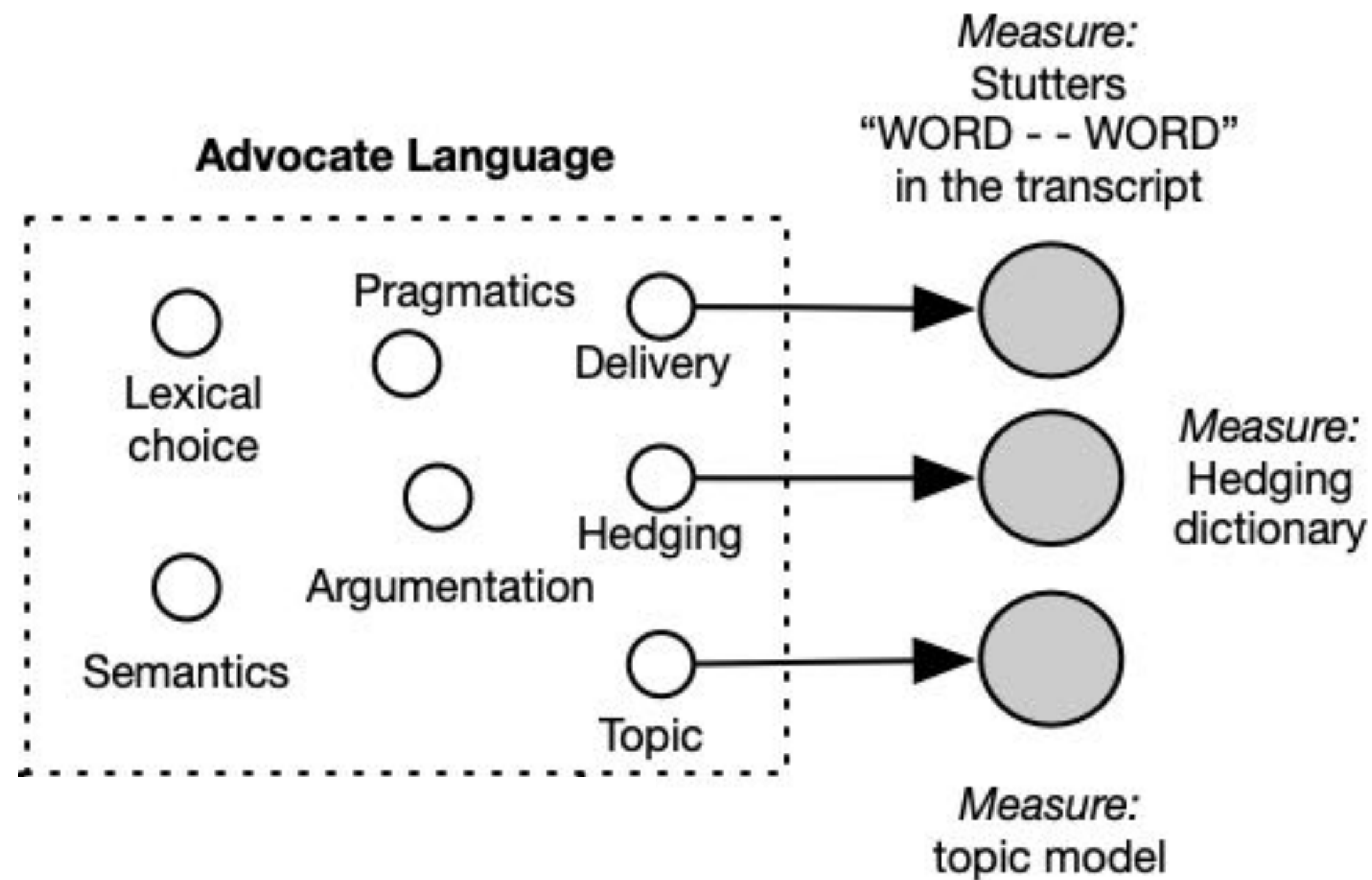
Recommendations

- Hypothetical manipulations

- Causally independent mediators

- Substantive theory

- Measurement error

Next steps

- Empirical estimates from real data
- Address causal dependence between temporal utterances
- Analyze between-judge and between-court temporal estimates

Thanks! Questions?