Type III and Type IV Errors:

Statistical Decision-Making Considerations in addition to Rejecting and Retaining the Null Hypothesis

Charlotte Ursula Tate, Ph.D.
San Francisco State University

Identifying Type III and IV Errors to Improve Science

- Behavioral science has become good at identifying factors related to Type I and II errors
- Zeitgeist in psychology is to avoid false positives and increase visibility of true negatives
- Type III and IV errors will help behavioral science create as stronger theory-method-statistics connection
    - Alongside, rather than supplanting Type I and II errors

Type I and Type II Errors Need Complementary Considerations

- Replication alone appears not to be able to tell us why two studies didn't replicate

Article 1

Research Article

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

The Fluctuating Female Vote: Politics, Religion, and the Ovulatory Cycle

Psychological Science
24(6) 1007–1016

DOI: 10.1177/0956797612466416
pss.sagepub.com

[redacted]

Article 2

Commentary

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Women Can Keep the Vote: No Evidence That Hormonal Changes During the Menstrual Cycle Impact Political and Religious Beliefs

Psychological Science
2014, Vol. 25(5) 1147–1149

DOI: 10.1177/0956797613520236
pss.sagepub.com
SAGE

[redacted]

The Lack of Conceptual Coverage

- Mosteller (1948) and Kimball (1957) (among others) noticed that there additional errors that one could make

- Type III Error was introduced to address this lack of coverage

- Type III Error had two foci:

  - H_0 considerations [specific]

  - Reasoning errors (outside H_0) that concerned model building—>statistical evaluation [generalized]

The Focus of Type I and II Errors

- Type I and II errors focus a researcher's attention on the "evaluation" aspect of the reasoning chain:

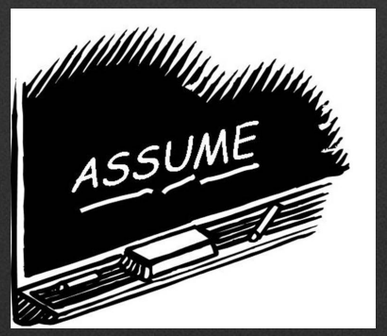
Model Building

Theory

Operationalization

Data Structure (planned)

?



Evaluation

Actual Data Structure (EDA)

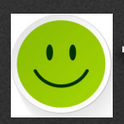
Statistical Tests Performed

Findings (e.g., NHST, effect size)

False positive

True positive

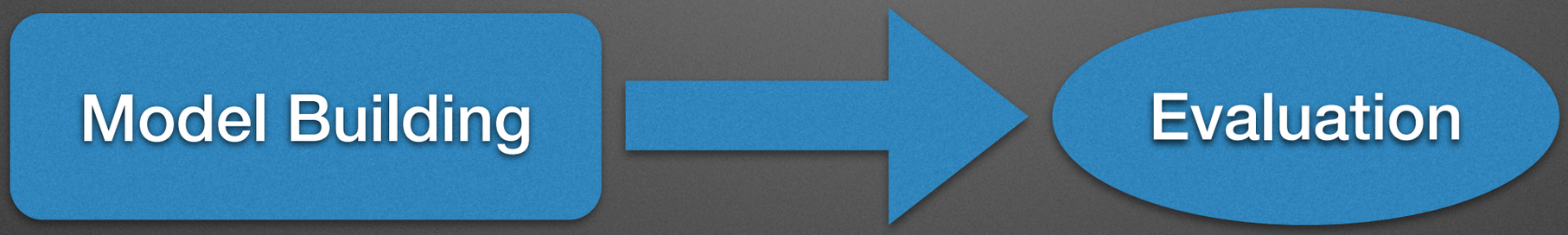
False negative



True negative

The Generalized Type III Error

- *Wrong model, right answer* focuses a researcher's attention on the two main parts of the reasoning chain:



Theory
Operationalization
Data Structure (planned)

Actual Data Structure (EDA)
Statistical Tests Performed
Findings (e.g., NHST, effect size)

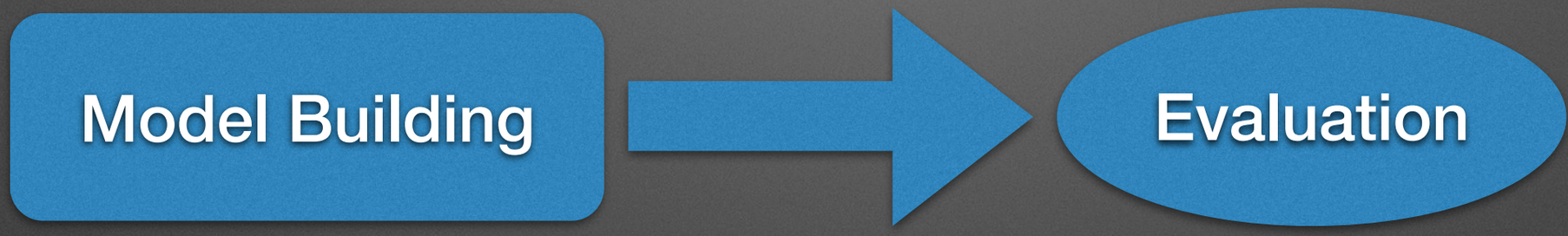
- Type III error is (graphically):

Some Influences on Generalized Type III Error

- Common Influences on Type III Error would be:
  - Incorrect operationalization of variables
  - Poor theory (e.g., ad hoc explanations of findings)
  - Mis-identifying causal architecture (Schwartz & Carpenter, 1999)
    - For instance, focusing on inter-individual factors for homelessness rather than structural factors
      - Confuses who becomes homeless with why homelessness exists

The Generalized Type IV Error

- *Right model, wrong answer* focuses a researcher's attention on the two main parts of the reasoning chain:



Theory
Operationalization
Data Structure (planned)

Actual Data Structure (EDA)
Statistical Tests Performed
Findings (e.g., NHST, effect size)

- Type IV error is (graphically):

Some Influences on Generalized Type IV Error

- Common Influences on Type IV Error would be:
  - Collinearity among predictors
  - Aggregation bias
  - Wrong test for data structure

One Example of Type IV Error

- Article 1 ovulation effects
- DVs: Religiosity, Imagined Political Behavior
- Predictors: Ovulation Status, Relationship Status, OS x RS

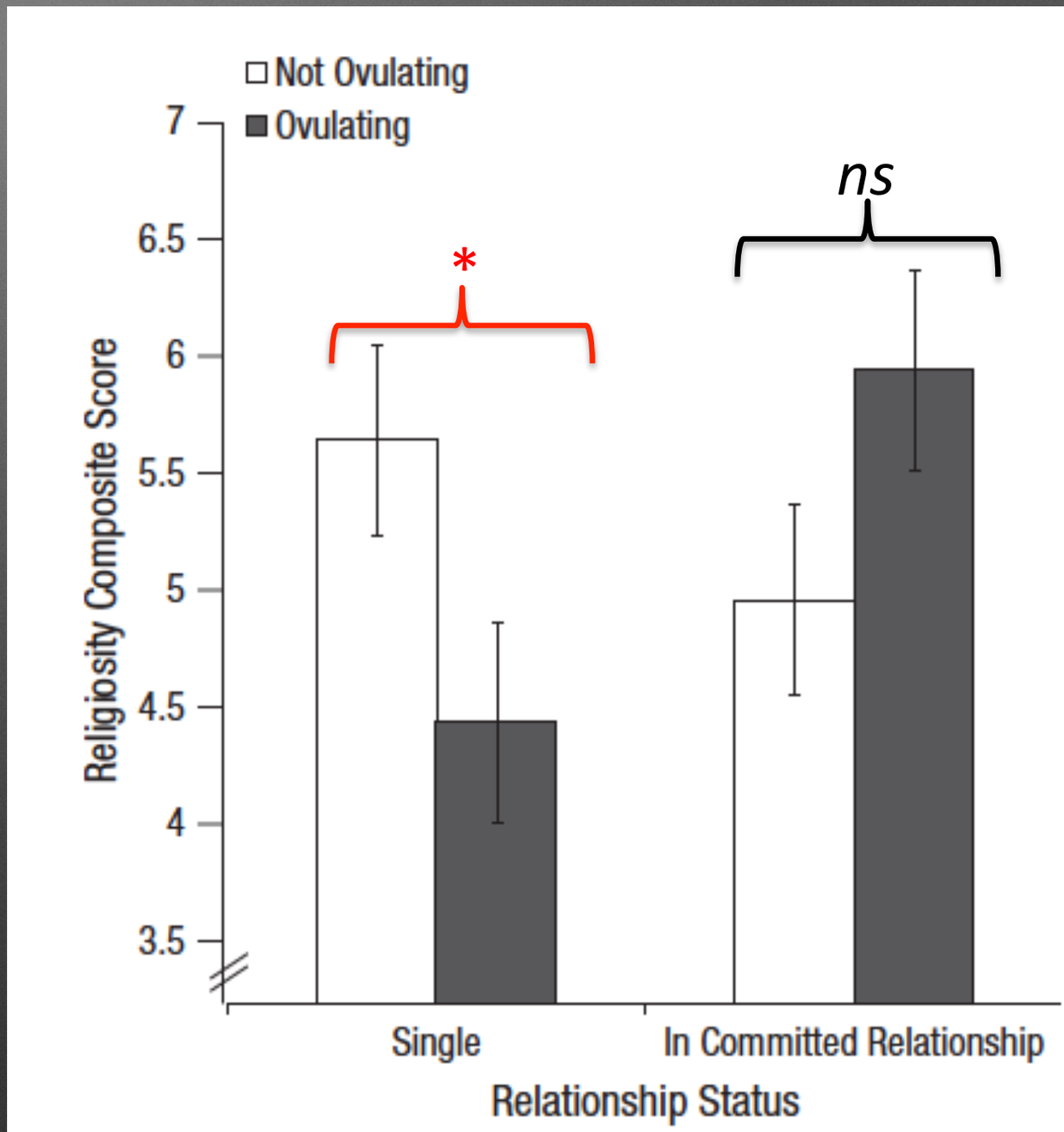


Fig. 1. Results from Study 1: women's religiosity as a function of fertility and relationship status. Higher scores indicate greater religiosity. Error bars represent standard errors of the mean.

Type IV Error Identified (Article 1)

Study 1

Demographic information. Women in relationships were older (mean age = 29.97 years) than were single women (mean age = 25.51 years), $p < .001$, and were more likely to have children (45.0%) than were single women (9.7%), $p < .001$. In the full sample, 70.2% of par-

5.5% were Asian American, 4%
Pacific Islander, 3.6% were His
American Indian/Native Americ
race or ethnicity, and 0.4% wer
nicity. Participants indicated the
hold income by selecting
descriptions: $15,000 or less, $1
$35,000, $35,001–$50,000, $5(
$100,000, $100,001–$150,000, o
two groups did not differ in e
income level ($p = .57$), and ethn
fertility.

Study 2

Demographic information. On average, women in relationships were older (mean age = 29.66 years) than were single women (mean age = 25.22 years), $p < .001$, and were more likely to have children (53.3%) than were single women ($M = 15.8\%$), $p < .001$. Participants

selecting one of the following descriptions: $15,000 or less, $15,001–$25,000, $25,001–$35,000, $35,001–$50,000, $50,001–$75,000, $75,001–$100,000, $100,001–$150,000, or $150,000 or more. Women in relationships had a higher annual income level (M = $35,001–$50,000 per year) than did single women (M = $15,001–$25,000 per

One Example of Type IV Error

- Article 1 ovulation effects
- DVs: Religiosity, Imagined Political Behavior
- Predictors: Ovulation Status, Relationship Status, OS x RS
- Collinearity creates: Relationship Status + Age + Child Status as *Relationship Status Moderator*
- ***Maybe*** Article 2 had **lower collinearity in the replication**

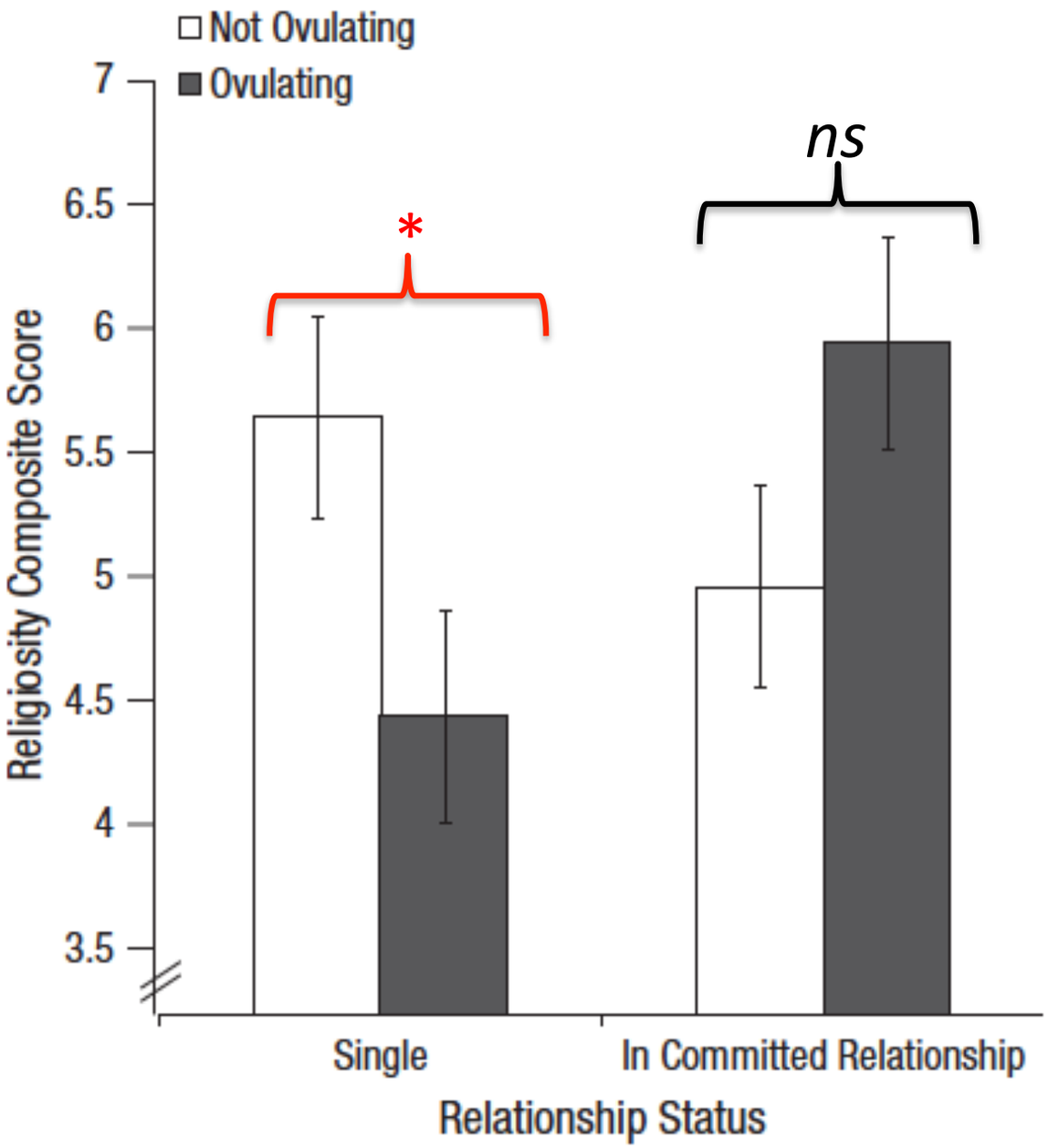


Fig. 1. Results from Study 1: women's religiosity as a function of fertility and relationship status. Higher scores indicate greater religiosity. Error bars represent standard errors of the mean.

The Import of Type III and IV Errors

- Focusing attention on the two main parts of the reasoning chain *specifies the location* of the error:

Model Building	Evaluation
Theory	Actual Data Structure (EDA)
Operationalization	Statistical Tests Performed
Data Structure (planned)	Findings (e.g., NHST, effect size)
Type III		
Type IV		

Reasoning Errors Unite!
(non-exhaustive list)

- TYPE III ERROR
- Mosteller's (1948) directional prediction error
- A. Gelman's (2000) Type S error
- (A. Gelman's (2000) Type M error?)
- Schwartz & Carpenter's Type III error
- Hyman-Tate (2015) theory-timing criterion for mediation analysis

- TYPE IV ERROR
- Rosenthal & Rosnow's (1986) cell mean error for simple effects
- Umesh et al. (1996) Type IV error for interaction terms
- A. Gelman's (2000) Type M error
- (A. Gelman's (2000) Type S error?)
- MacKinnon & Pirlott's (2015) confounder bias on mediators