

第四讲 实验室实验

张开平 助理教授

清华大学社会科学学院

2021年3月15日

实验室实验的基本要素

- 人为设置的场景 artificial settings
- 对自变量进行干预 manipulation of independent variable
- 随机化 randomization
- 严格控制的流程 controlled conditions

实验室实验案例

- Ash Conformity Experiment
 - 体现了实验室实验的哪些要素？
 - 它的研究问题和因果模型是什么？
 - 你能否想出其他的实验设计？
- The Monkey Business Illusion

The Door Study

基于实验的理论构建

- 检验并得出“放之四海而皆准”的一般性规律
 - 不局限于时间、地点和人群 not subject to context and population
 - 对不同人群的效果一样吗 (heterogeneous treatment effect)
 - 通过什么机制形成 (through what mechanism)

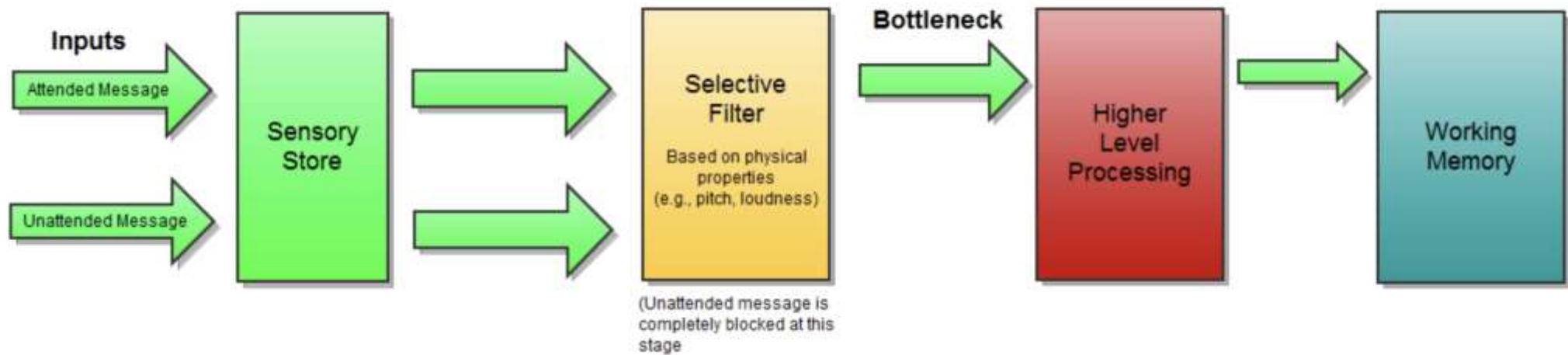
What causes Y?

What are the dimension of X?

By what mechanisms do these effects occur?

Under what conditions do these effects occur?

Broadbent's Filter Model



什么研究适合做实验室实验

- 观察性研究的问题

- 自我汇报偏差
- 内生性问题
- 无法证实某个因素的因果作用
- 基于观察性数据与基于实验的研究可能得到相反的效果

- 政治学研究中实验方法的优势

- 严格的因果推断
- 隔离其他因素，检验某个因素的因果作用
 - 例如：竞选宣传中的信息、候选人、渠道、语调、表情手势等各方面因素，每个因素的单独影响
- 能够复制和进行元分析
- 进行理论构建

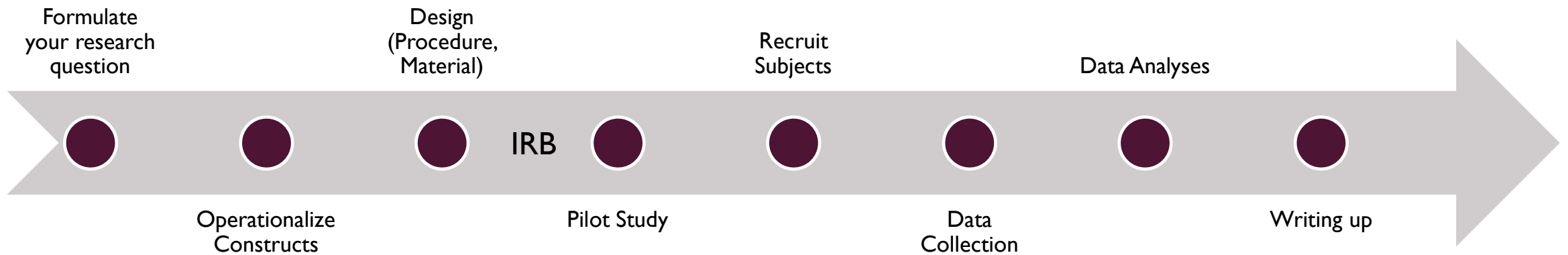


观看竞选广告是否对鼓励投票有正向作用？

实验室实验的局限

- 问题 External Validity/Generalizability
 - 实验室的人造环境多大程度上能推广至其他情境？
 - 基于学生样本的结论多大程度上能推广至普通公众？
- 解决方法
 - 使用其他方法验证
 - 在具有代表性的人群做实验

实验室实验的基本流程



概念操作化 OPERATIONALIZATION OF CONCEPTS

- The formal procedure that links scientific concepts to data collection
 - Concept → construct that is distinguishable, measurable, understandable by empirical observation
 - 优先使用成熟量表
 - 练习
 - 健康
 - 情绪
 - 能力
 - 排他主义
 - 民族主义
 - 威权人格

实验干预设计

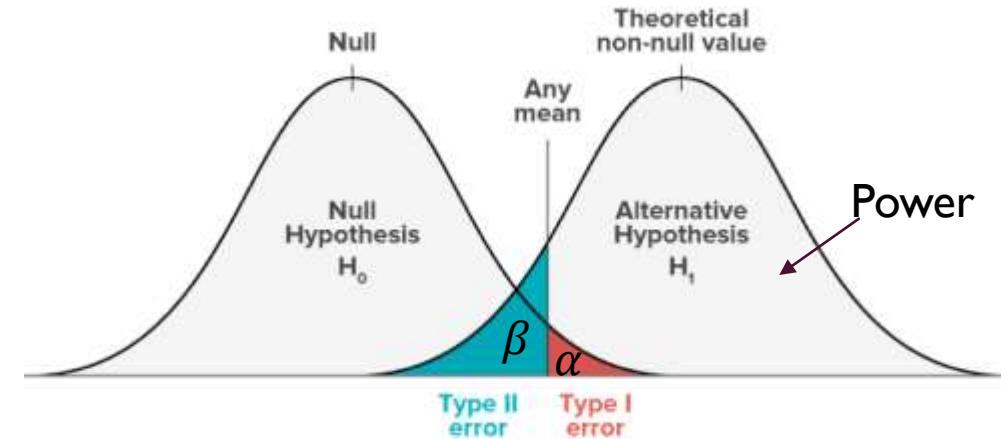
- 文字、图片、视频、场景、互动、任务、虚拟现实
 - [Daisy ad](#)
 - [Still Face Experiment](#)
 - [Fire Escape](#)
 - Pre-test
 - As strong as possible!
- Look into literature and be creative!

实验流程设计

- 流程标准化
 - 每个被试实验开始时，都由RA介绍同一段话，读相同的材料，回答相同的问题
- 减少实验者效应 (demand effect)
 - 暂时隐瞒实验的真实目的
 - 使用不突兀的效果测量
 - 增加实验场景的真实性
- Within vs. Between Subjects
- 如何施加实验干预
- How many conditions?
- 在何种人群上做实验,如何招募？需要多少样本？

Calculating Sample Size with Power Analysis

- Power: the probability of correctly rejecting a false null hypothesis (H_0)
 - If your power for detecting the true mean difference is 40%, then you will have a 60% chance of getting a $p > 0.05$.
 - Power > 0.8
 - The bigger the sample size, the stronger the power.
 - Done before experiment!
- Power analysis: estimate an appropriate sample size
 - Too big: a waste of resource
 - Too small: miss statistical significance
 - Relate to cost and ethnicity: justification for IRB and Grant.



被试招募和样本量计算

- 广告、激励、滚雪球.....
- 计算需要多少样本
 - 假设效果存在，样本中达到统计显著所需要的样本量 (N)
 - 给定每个实验组的均值和标准差，给定需要的 β 和 α 水平
 - 例如，两组间独立样本T检验，我们可以根据以下公式计算每个组所需要的样本量

$$n = \frac{2s^2 (z_{\beta} + z_{\alpha/2})^2}{\bar{y}_1 - \bar{y}_2}$$

- a. The denominator represents the effect size
- b. s^2 = the pooled variance across conditions
- c. The z-score for the conventional β of 0.80 is 0.84
- d. The z-score for the conventional two-tailed α (0.05/2) is 1.96

POWER ANALYSIS

- **power** in Stata; **pwr** package in R

- 如何决定mean和SD? -> 以前的相似题目研究、前测

Between-subjects

```
. power twomeans 4 5, sd1(2) sd2(2)
```

Performing iteration ...

Estimated sample sizes for a two-sample means test
Satterthwaite's t test assuming unequal variances
Ho: $m_2 = m_1$ versus Ha: $m_2 \neq m_1$

Study parameters:

```
alpha = 0.0500
power = 0.8000
delta = 1.0000
m1 = 4.0000
m2 = 5.0000
sd1 = 2.0000
sd2 = 2.0000
```

Estimated sample sizes:

```
N = 128
N per group = 64
```

Within subjects
(repeated measures)

```
. power onemean 4 5, sd(2)
```

Performing iteration ...

Estimated sample size for a one-sample mean test
t test
Ho: $m = m_0$ versus Ha: $m \neq m_0$

Study parameters:

```
alpha = 0.0500
power = 0.8000
delta = 0.5000
m0 = 4.0000
ma = 5.0000
sd = 2.0000
```

Estimated sample size:

```
N = 34
```

POWER ANALYSIS IN R

```
install.packages("pwr")  
library(pwr)
```

```
pwr.t.test(d = , sig.level = , power = ,  
           type = c("two.sample", "one.sample", "paired"))
```

```
> pwr.t.test(d = (5-4)/2, power = .8, sig.level = .05, type = "two.sample")
```

Two-sample t test power calculation

Per group →

```
      n = 63.76561  
      d = 0.5  
sig.level = 0.05  
  power = 0.8  
alternative = two.sided
```

NOTE: n is number in *each* group

```
> pwr.t.test(d = (5-4)/2, power = .8, sig.level = .05, type = "one.sample")
```

One-sample t test power calculation

```
      n = 33.36713  
      d = 0.5  
sig.level = 0.05  
  power = 0.8  
alternative = two.sided
```

Manipulation Check

- Whether the treatment manipulated the independent variable (IV) as intended; whether treatment successfully induced variance in the IV.
- 检查实验干预是否成功与汇报实验结果同样重要！
 - 如果假设得不到证实，并不知道是因为假设不成立，还是干预不成功 (Type II error)
 - 研究者常低估产生干预效应所需要的干预强度
 - “For experiments to have the best chance of succeeding, the researcher needs to ensure that the manipulation of the IV is **as strong as possible**. Indeed, if there were **a first rule of experimentation**, this might be it.” (*Sage Handbook of Methods in Social Psychology*)
 - Avoid weak or ineffective manipulation! Run pre-test.
 - 实验干预可能产生相反效果或实验计划以外的效果
 - 政治学常常采用间接方式干预一个潜在变量，因此评估 **construct validity** 至关重要，即确认实验干预是否实际上干预了因变量
 - 调查实验、网络实验无法保证人们的注意力，更加需要干预检验 (**inattentive respondents problem**)
 - 方法：free recall, ask questions “你注意到xx了么？” “你现在的心情是怎样的？” 观测其在页面停留时间等

Randomization checks/Balance Testing

- Aa table of the distribution of pre-treatment measures across treatment groups.
 - “Randomization check shows that demographics and political predispositions do not jointly predict treatment assignment”

Covariates Inclusion?

- 对协变量的引入需谨慎并基于理论，在实验干预之前测量（例如：政治知识）
- 当协变量被引入回归分析，干预效应估计会相应改变。因此需要首先展示不含有协变量的实际干预效应。
“75% of experimental results never show the reader the dependent variable means by experimental condition or a regression only including treatment effects.” (Gerber et al.,2014)

Multi-level Analysis

- When individuals are nested within groups, hierarchical structured data observations violate OLS heteroscedastic assumption. Multi-level analysis (or so-called Hierarchical Linear Model) should be used.

实验数据分析

- 实验设计超过两组时

- ANOVA （只能得出几组之间有区别，但不知道哪两组间有区别）
- Regression （得出某组对比参照组baseline的区别是否有区别）

```
anova      warmscale sexocond

          Number of obs =      555      R-squared      = 0.0064
          Root MSE      = 1.12441      Adj R-squared = 0.0028

-----+-----
Source | Partial SS      df      MS      F      Prob>F
-----+-----
Model | 4.5060811      2      2.2530405      1.78      0.1693
sexocond | 4.5060811      2      2.2530405      1.78      0.1693
Residual | 697.89257      552      1.2642981
-----+-----
Total | 702.39865      554      1.2678676
```

```
regress      warmscale i.sexocond

-----+-----
warmscale |      Coef.      Std. Err.      t      P>|t|      [95% Conf. Interval]
-----+-----
sexocond |
Bisexual |      -0.170      0.117      -1.456      0.146      -0.400      0.059
Gay/Lesbian |      -0.207      0.117      -1.769      0.078      -0.436      0.023
_cons |      4.889      0.083      59.142      0.000      4.727      5.052
-----+-----
```

FACTORIAL DESIGN

```
. mtable,          at(gencond=(0 1) sexocond=(1 2 3))
```

Expression: Linear prediction, predict()

	gencond	sexocond	xb
1	0	1	4.412
2	0	2	4.718
3	0	3	4.997
4	1	1	5.382
5	1	2	4.720
6	1	3	4.357

```
<- Heterosexual Men
<- Bisexual Men
<- Gay Men
<- Heterosexual Women
<- Bisexual Women
<- Lesbian Women
```

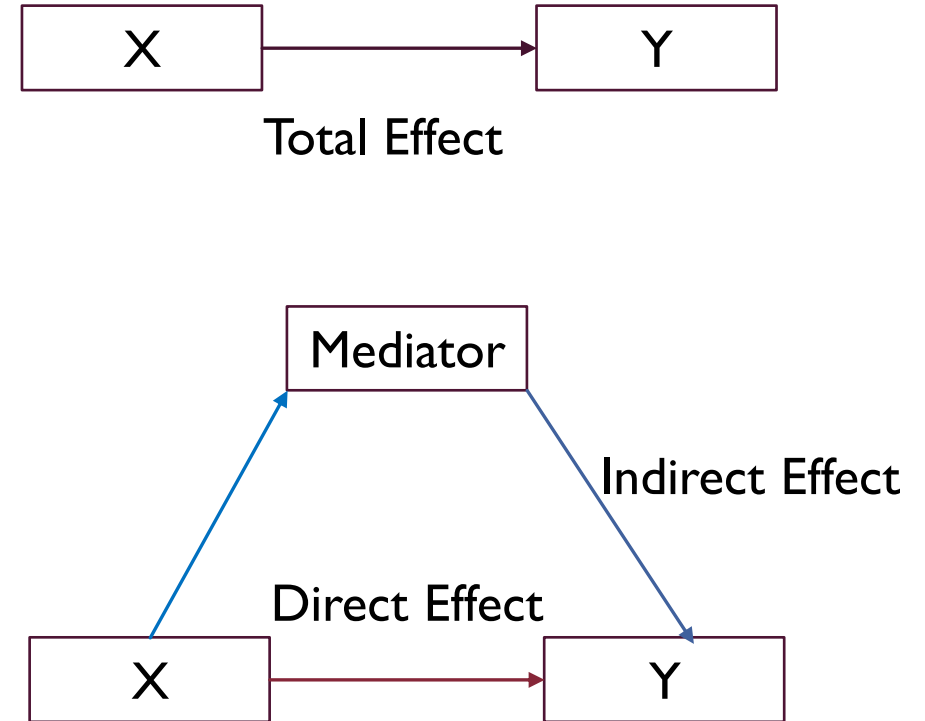
```
. regress          warmscale i.gencond##i.sexocond
```

Source	SS	df	MS	Number of obs	=	555
Model	66.9292922	5	13.3858584	F(5, 549)	=	11.56
Residual	635.469356	549	1.15750338	Prob > F	=	0.0000
				R-squared	=	0.0953
				Adj R-squared	=	0.0870
				Root MSE	=	1.0759
Total	702.398649	554	1.2678676			

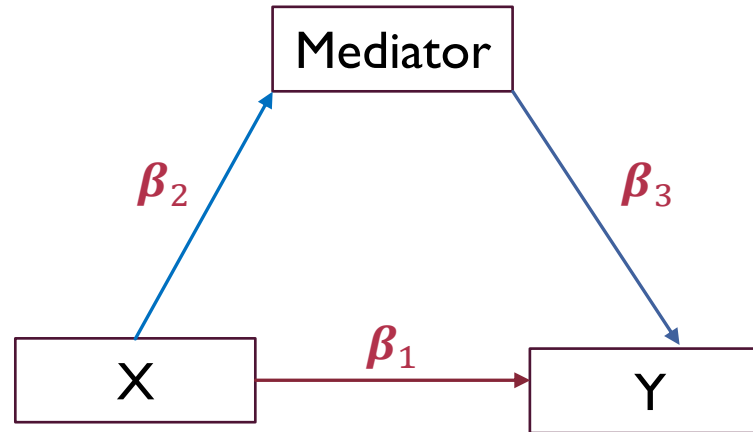
warmscale	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
gencond						
Women	0.970	0.158	6.128	0.000	0.659	1.280
sexocond						
Bisexual	0.306	0.157	1.949	0.052	-0.002	0.614
Gay/Lesbian	0.585	0.157	3.728	0.000	0.277	0.893
gencond#						
sexocond						
Women #						
Bisexual	-0.968	0.224	-4.326	0.000	-1.407	-0.528
Women #						
Gay/Lesbian	-1.610	0.224	-7.195	0.000	-2.049	-1.170
_cons	4.412	0.111	39.761	0.000	4.194	4.630

Mediation 中介效应

- **Explain why or how X affects Y .**
 - 如果发现观看竞选广告对于鼓励投票有作用，那么为什么观看竞选广告有作用？
 - Does physical abuse in early childhood lead to deviant processing of social information that leads to aggressive behavior?
 - Does trauma affect brain stem activation in a way that inhibits memory?
 - If an intervention increases secure attachment among young children, do behavioral problems decrease when the children enter school?



Mediation Analyses



$$\text{Step 1: } Y = \beta_0 + \beta_1 X + \varepsilon$$

$$\text{Step 2: } Y = \beta_0 + \beta_2 X + \varepsilon$$

$$\text{Step 3: } Y = \beta_0 + \beta_4 X + \beta_3 M + \varepsilon$$

Mediation occurs when 1) there is a statistically significant indirect effect and 2) the direct effect is smaller than the total effect.

Mediation Analyses

- *Mediation* package in *R* (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014).
- Sobel test in *Multilevel* package and bootstrapping in *Mediation* package for significance test

```
library(mediation)
results <- mediate(model.M, model.Y, treat='X', mediator='M',
                  boot=TRUE, sims=500)
summary(results)
```

#	Estimate	95% CI Lower	95% CI Upper	p-value
# ACME	0.3565	0.2155	0.5291	0.00
# ADE	0.0396	-0.1761	0.2598	0.66
# Total Effect	0.3961	0.1563	0.5794	0.00
# Prop. Mediated	0.9000	0.5254	1.8820	0.00

β_4
 β_1

```
### ACME = 0.3565, 95% CI [0.2155, 0.5291] # significant!
### ACME stands for Average Causal Mediation Effects
### ADE stands for Average Direct Effects
### Total Effect is a sum of a mediation (indirect) effect and a direct effect
```

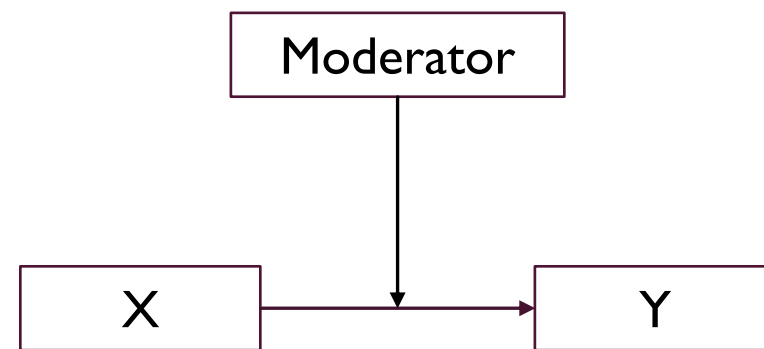
$$\beta_1 = \beta_4 + \beta_2 * \beta_3$$

The goal of mediation analysis is to obtain the indirect effect and see if it is statistically significant.

Mediation vs. Moderation

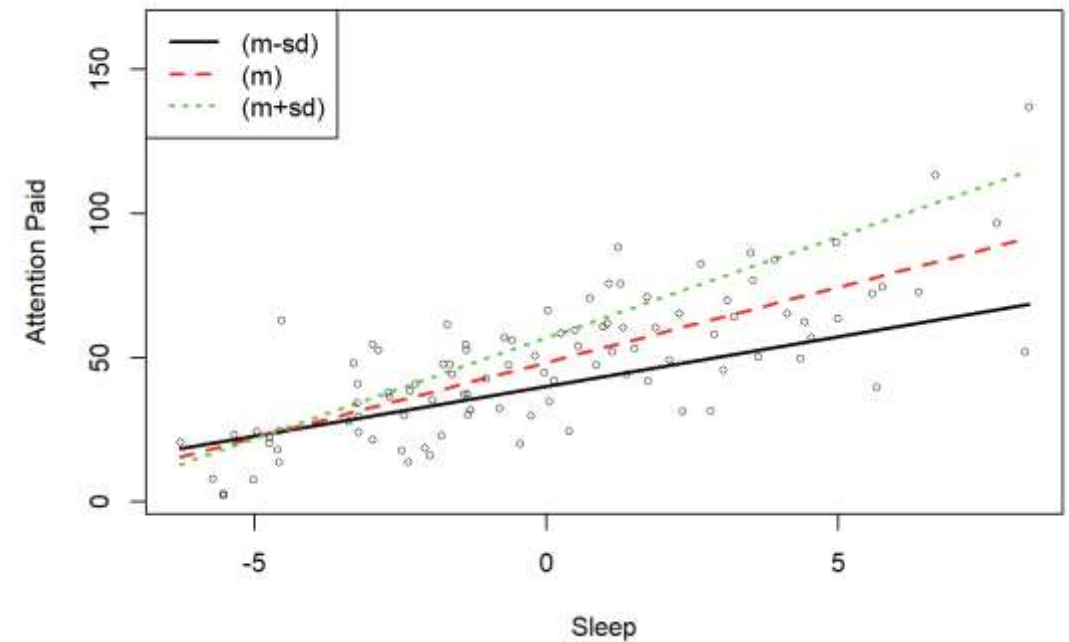
■ Moderation 调节效应

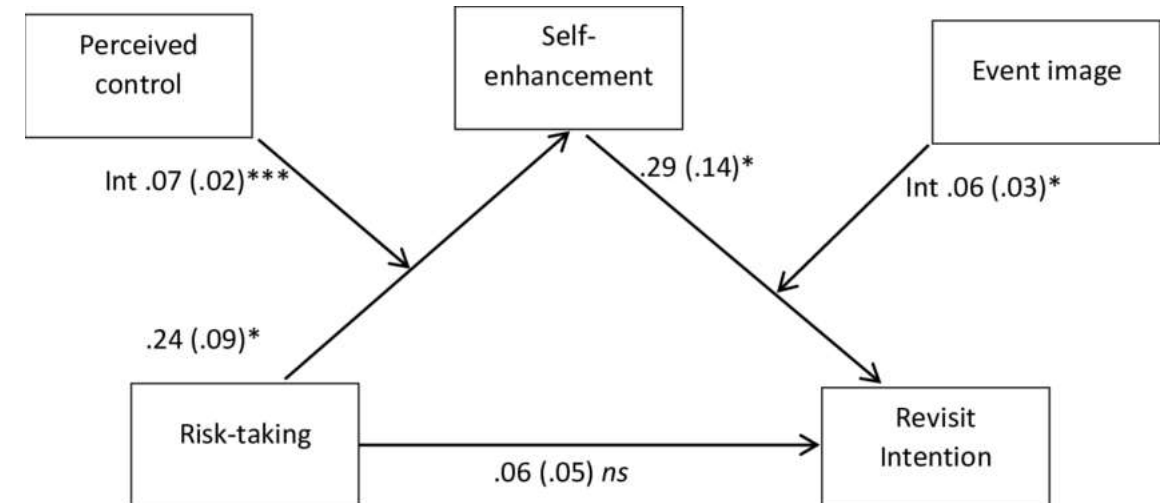
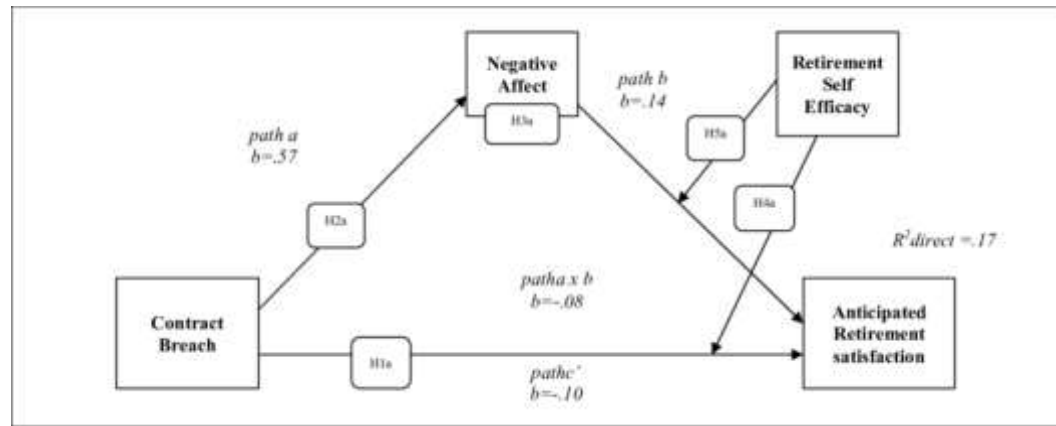
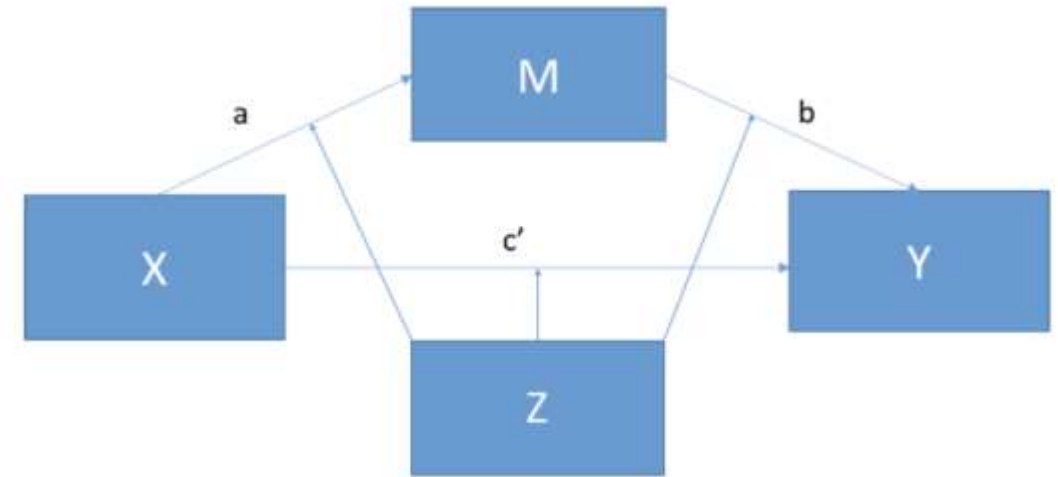
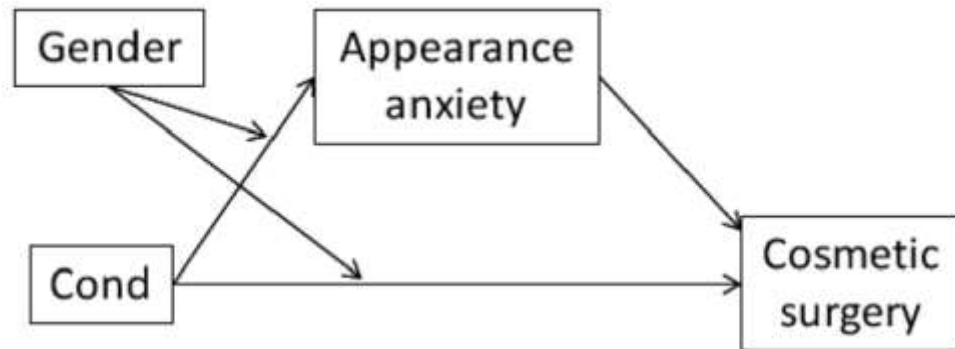
- when or under what conditions an effect occurs
- Moderators can strengthen, weaken, or reverse the direction of a relationship.
- 比如：新冠病毒对人体的破坏程度受年龄的影响，年龄越大，破坏力越大



Moderating Analyses

- Interaction term: $X*M$
- Choose your X and Moderator based on theories.
- it is important to **mean center** both your moderator and your IV to reduce multicollinearity and make interpretation easier if they are continuous variables.

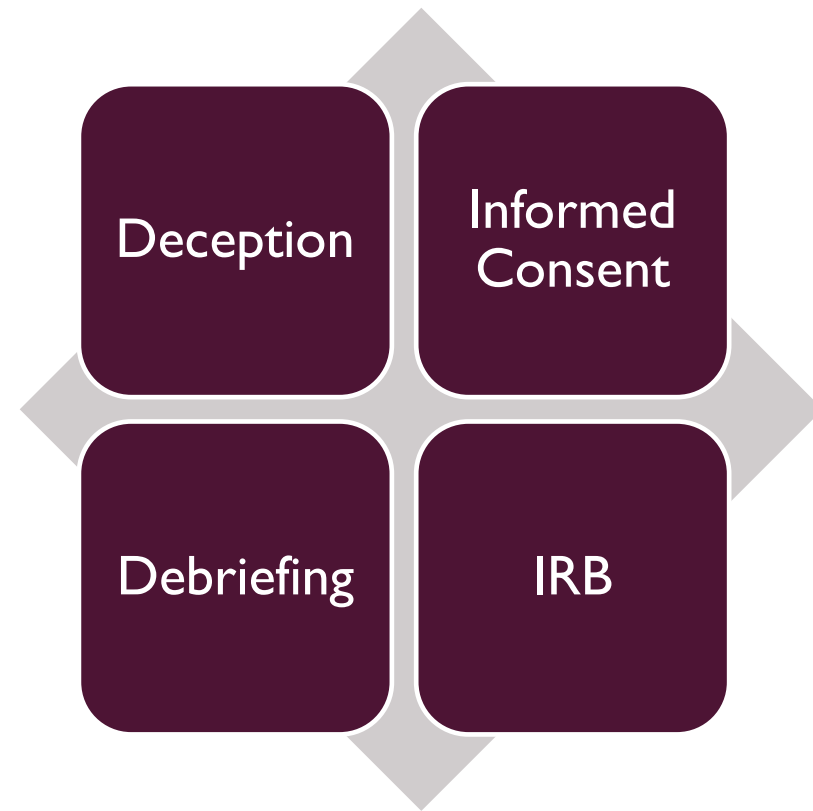




提升实验室实验的外部效度

- 招募一般人群被试
- 使用与现实中相似的实验材料
- 实验室模仿真实场景
- 叠加多个实验：实验室实验 + 实地实验
- 使用其他方法例如问卷数据印证

实验伦理

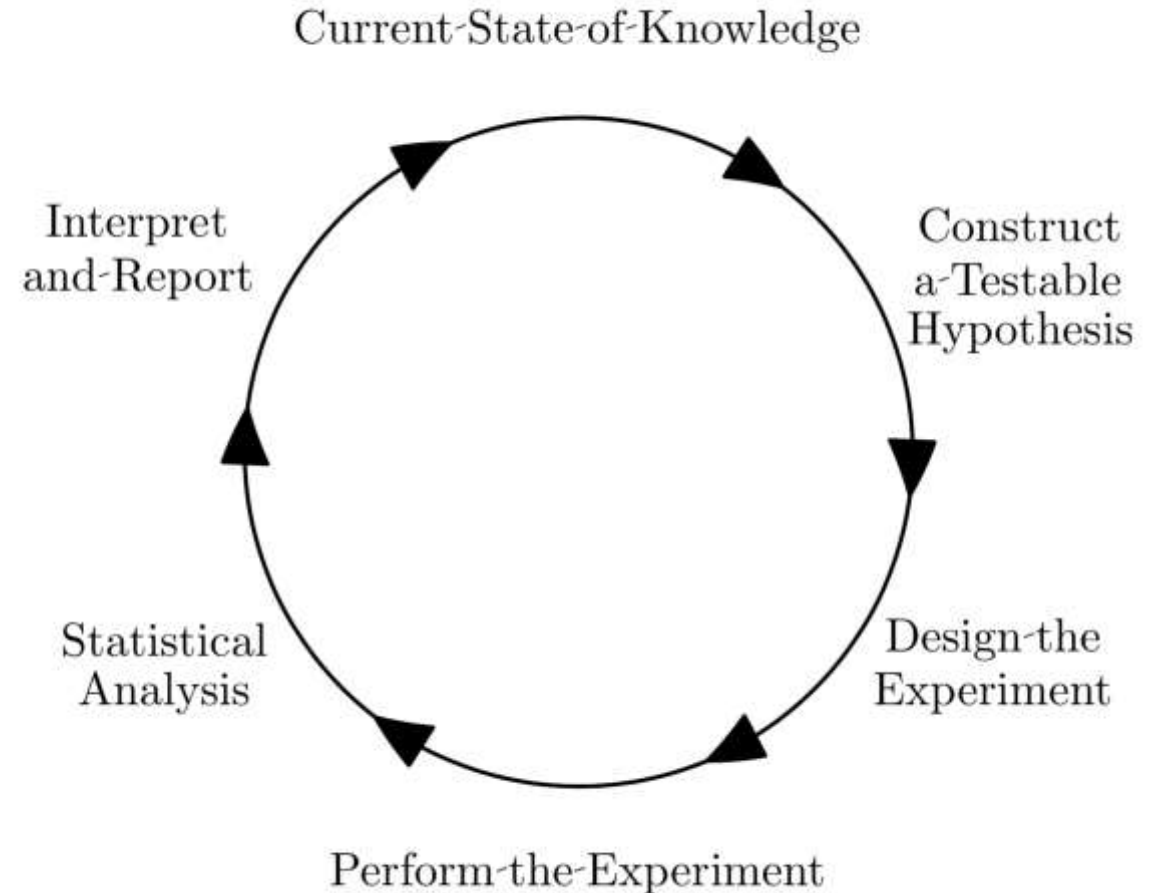




解剖一只麻雀： Gilliam & Iyengar study

Formulate Research Question

- What is the research question?
 - “Our objective in this article is to evaluate the relative contribution of each element of the crime news script – the focus on violent crime and the inclusion of racial imagery – on public opinion.” (Gilliam & Iyengar, 2000)
- Construct hypotheses
 - What is the causal model?



Experimental Design

Crime Violence
+ Black Male

Crime Violence
+ White Male

Crime Violence

No Crime News



推荐阅读

- Cohen, J. (1992). Statistical power analysis. *Current directions in psychological science*, 1(3), 98-101.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Iyengar, S. (2011). Laboratory experiments in political science. *Cambridge handbook of experimental political science*, 73-88.
- Gilliam Jr, F. D., & Iyengar, S. (2000). Prime suspects: The influence of local television news on the viewing public. *American Journal of Political Science*, 560-573.
- Mutz, D. C., & Pemantle, R. (2015). Standards for experimental research: Encouraging a better understanding of experimental methods. *Journal of Experimental Political Science*, 2(2), 192-215.
- Sansone, C., Morf, C. C., & Panter, A. T. (Eds.). (2003). *The Sage handbook of methods in social psychology*. Sage Publications.
- Seltman, H. J. (2012). *Experimental design and analysis*.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). *Mediation: R package for causal mediation analysis*.