瑞妮塔·科尔曼

.•.



为社会科学设计实验Designing

**为社会科学设计实验**

##### S Shao翻译

Sara Miller McCune于1965年创立SAGE出版社，旨在支持有用知识的传播并教育全球社群。SAGE每年出版1000多种期刊和800多本新书，涵盖广泛的学科领域。我们不断增加的图书馆产品选择包括档案、数据、案例研究和视频。SAGE仍然由我们的创始人多数持有，在她的有生之年之后，将由一个慈善信托所有，确保公司的持续独立性。

洛杉矶 | 伦敦 | 新德里 | 新加坡 | 华盛顿特区 | 墨尔本

**为社会科学设计实验**

如何规划、创建和执行使用实验进行研究

### 里尼塔·科尔曼

*德克萨斯大学奥斯汀分校*





提供信息：

SAGE出版社股份有限公司

地址：2455 Teller Road

千橡市，加利福尼亚州91320

电子邮件：order@sagepub.com

SAGE出版社有限公司

地址：1 Oliver's Yard

伦敦市，55 City Road

EC1Y 1SP，英国

SAGE出版社印度私人有限公司

地址：新德里莫汉合作工业区B 1/I 1

Mathura Road

110 044，印度

版权所有 © 2019 SAGE Publications, Inc.

ISBN 978-1-5063-7732-2

Copyright  2019 by SAGE Publications, Inc.

保留所有权利。未经出版社书面许可，本书任何部分均不得以任何形式或任何方式进行复制或利用，包括复印、录制或任何信息存储和检索系统。

在美国印刷

ISBN 978-1-5063-77

SAGE出版社亚太区私人有限公司 阅读-非商业用途（2023）

3 Church Street

#10-04 Samsung Hub

新加坡 049483

收购编辑：海伦·萨尔蒙

开发编辑：切尔西·尼夫

编辑助理：梅根·奥海弗南

制作编辑：奥利维亚·韦伯-斯特尼斯

校对编辑：黛安·温赖特

排版人员：C&M Digitals Ltd.

校对员：詹妮弗·格鲁巴

索引制作人：让·卡萨莱诺

封面设计师：金坎·西亚姆

市场经理：苏珊娜·戈尔德斯

“这本书提供了许多示例和实践指导，可以帮助学生学习如何进行实验。”

- 弗朗西斯·O·阿德奥拉（Francis O. Adeola），新奥尔良大学

“学习实验设计必备的一本书籍——从构建研究问题到解释结果，本书全面涵盖。”

- 莎拉·埃尔·赛义德（Sarah El Sayed），德克萨斯大学阿灵顿分校

“本书为学生提供了对实验方法论的优秀解释：不仅仅是对实验方法的元素进行描述，还包括对这些元素的起源、它们如何以及为什么相互配合以及学生如何成为学者的理解。”

- 凯尔·J·霍洛迪（Kyle J. Holody），海岸卡罗来纳大学

“这是一本内容全面、清晰生动并包含许多优秀例子的实验教材。Renita Coleman 包含了学生在发展、设计和执行方法上健全的实验所需的一切。”

- 格伦·莱什纳（Glenn Leshner），俄克拉荷马大学

“本教材详细且组织良好，每个主题都以逻辑方式逐步建立在基础知识之上。学生将发现很容易掌握研究概念并通过提供的练习和例子进行实践。‘学习聚光灯’场景和实际例子是展示研究设计的有效工具，‘测试你的知识’问答则强化学生的学习。”

- 珍妮特·里德-赫克特（Janet Reid-Hector），罗格斯大学

“本书写作精细且易读。如果您需要掌握实验设计的概念，本书将指导您了解实验设计的内容以及在开展研究时需要牢记的事项。”

- 饶智勋（Ji Hoon Ryoo），弗吉尼亚大学

前言 xv

关于作者 xxi

|  |  |  |
| --- | --- | --- |
| 章节一 | * 发现因果关系 | 1 |
| 章节二 | * 伦理和历史上著名的实验 | 23 |
| 章节三 | * 理论，文献和假设 | 55 |
| 章节四 | * 实验的类型 | 89 |
| 章节五 | * 内部效度和外部效度 | 111 |
| 章节六 | * 因子设计 | 145 |
| 章节七 | * 随机分配 | 173 |
| 章节八  抽样和效应大小 211 | | |
| 章节九 | * 刺激和操作检查 | 249 |
| 章节十 | * 工具和测量 | 289 |
| 章节十一 | * 伦理审查委员会和进行道德实验 |  |
|  |  | 337 |
| 术语表 |  | 369 |
| 索引 |  | 377 |

For Review-No Commercial Use(2023)

[前言 xv](#_TOC_250030)

[致谢 xix](#_TOC_250029)

[关于作者 xxi](#_TOC_250028)

1. 发现因果关系 1

[因果关系 2](#_TOC_250027)

[实验与其他方法的比较 3](#_TOC_250026)

[实验的基本标准 4](#_TOC_250025)

[实验要素 6](#_TOC_250024)

[变异性 6](#_TOC_250023)

[混杂因素 7](#_TOC_250022)

[对照组 9](#_TOC_250021)

分配 9

[开始你自己的研究 11](#_TOC_250020)

[撰写问题陈述 12](#_TOC_250019)

[回答“那又怎样”的问题 14](#_TOC_250018)

1. 历史上的伦理和著名实验 23

[坏血病研究 24](#_TOC_250017)

[查尔斯·皮尔斯的贡献 27](#_TOC_250016)

[罗纳德·费舍尔的绘图和茶叶 28](#_TOC_250015)

[B·F·斯金纳：小样本，高科技 30](#_TOC_250014)

[斯坦利·米尔格拉姆震撼世界 33](#_TOC_250013)

[菲利普·津巴多：在斯坦福地下室引发良知 40](#_TOC_250012)

[结论 44](#_TOC_250011)

1. 理论、文献和假设 55

[文献综述 58](#_TOC_250010)

[关于撰写实验文献综述的建议 60](#_TOC_250009)

它不是一篇书评60

不要成为维基百科61

做出理论贡献 63

连接各个点63

[概念定义与操作化 65](#_TOC_250008)

[涉及多个实验的文献综述 66](#_TOC_250007)

[假设和研究问题 68](#_TOC_250006)

[假设的基本原理 72](#_TOC_250005)

零假设与备择假设72

差异与方向73

[假设写作公式 75](#_TOC_250004)

[涉及多个自变量的假设 77](#_TOC_250003)

[研究问题 81](#_TOC_250002)

1. 实验的类型 89

Campbell和Stanley的实验分类 90

[三种前实验设计 91](#_TOC_250001)

单次案例研究91

单组前后测试设计91

静态组比较93

[三种真实实验设计 93](#_TOC_250000)

前后测试对照组设计93

Solomon四组设计94

仅后测试对照组设计94

##### 准实验 96

自然实验 100

场景实验 102

1. 内部和外部效度 **111**

生态效度和外部效度 112

泛化能力 113

随机抽样 **116**

两步随机化117

非响应偏差117

代表性 **118**

因果关系 122

逻辑推断 123

复制 124

自我复制 **124**

精确复制与概念复制 **125**

多个实验 **126**

外部复制 **127**

三角法 **129**

内部效度 129

三个基本标准 **130**

随机分配 **131**

混杂因素 **131**

历史因素132

成熟效应133

测试效应133

仪器效应134

统计回归135

选择效应135

退学效应135

1. 阶乘设计 **145**

单因素设计 146

阶乘设计 146

主效应和交互效应 **147**

阶乘符号表示法 **150**

设计表 **154**

选择设计 **156**

受试者在设计中的使用方式 158

组间设计 **158**

组内设计**Within-Subjects Designs 159**

混合阶乘设计 **161**

不完整因素设计（2023年） **162**

对照组 163

无干预对照组 **165**

对照组中的创造性 **166**

现状对照组 **166**

无对照组 **167**

1. 随机分配 **173**

随机分配的目的 174

避免混淆 **175**

随机是什么？ **176**

操作化随机分配 178

计算机随机化 **178**

调查实验 **182**

现场随机化 **183**

报告随机分配 183

平衡和不平衡设计 183

检查随机分配是否有效 184

聚合层次的随机分配 **185**

报告随机分配结果 **185**

当随机分配失败时 **187**

阻断、匹配和其他策略 189

阻断与简单随机分配 **190**

分层随机分配 **191**

其他事物的随机分配 191

交叉平衡设计 **194**

拉丁方设计 **194**

随机分配的阻力 196

1. 抽样和效应大小 **211**

学生样本 212

高于平均认知能力 **213**

服从权威 **217**

较弱的自我意识 **217**

其他众多差异 **217**

同质性 **218**

基本心理过程 **218**

保护学生免受偏见 **219**

亚马逊的在线众包平台 221

MTurk是否具有代表性？ **222**

优点和缺点 **222**

其他研究对象来源

225

招募 226

激励措施 228

样本大小和统计功效 229

效应大小 **231**

统计功效分析 **234**

观察到的统计功效237

1. 刺激和操作性检查 **249**

刺激的例子 251

刺激的创建建议 257

保持真实 **257**

尽量控制 **258**

仅变化研究对象 **259**

最大化比较 **260**

运用信息差异 **261**

需要创建多少刺激？263

固定因素与随机因素263

操作性检查 264

操作性检查、预测性检验还是试验研究？ **265**

直接与间接操纵 **266**

中介变量和心理状态 **266**

在调查实验中的操作性检查 **271**

实际问题 **272**

报告刺激和操作性检查 274

刺激描述 **274**

操作性检查描述 **276**

1. 仪器和测量 **289**

仪器 290

问卷调查 **290**

单个项目与指标292

优点和缺点293

隐蔽和非反应性测量 **297**

技术仪器 **297**

功能性磁共振成像299

心率301

皮肤电导301

其他生理仪器301

眼动跟踪302

屏幕注视303

网络跟踪303

延迟响应

304

次要反应任务305

连续反应工具306

思维列表和大声思考307

观察 **309**

观察者间的一致性310

其他隐蔽仪器 **312**

隐蔽问题 **312**

测量问题 312

测量层次 **315**

响应选择 **315**

顶部效应和底部效应316

构念效度 **317**

1. 机构审查委员会

道德实验 **337**

机构审查委员会 338

实验中的道德问题 342

欺骗 **342**

伤害 **344**

滥用权力 **345**

随机分配 **346**

保护受试者 346

知情同意 **347**

隐私和保密性 **348**

研究人员问题 349

数据伦理 **349**

抄袭和自我抄袭 **351**

作者资格问题 **352**

试点研究 354

术语表 369

索引 377

## 前言

四十年后，唐·坎贝尔和朱利安·斯坦利说：“这句话今天仍然是正确的。”2

五十多年过去了，我依然认为这句话仍然是真实的，特别是对于社会科学而言。

本书讲述的是如何获取足够和适当的数据。

它与大多数其他实验设计书籍有两个不同之处：首先，它专注于实验设计过程中的方法问题，而不仅仅是在数据收集后进行各种统计分析。从一开始就仔细设计研究是良好研究的关键。这在实验中尤为重要。本书关注的是研究人员在实验的程序性方面做出决策时的阶段，即在干预和处理之前。

本书将帮助读者逐步了解如何从一开始规划和执行实验，包括决定是否使用单因素或多因素设计、如何分配研究对象到组别、选择和收集样本、创建刺激和工具、进行操作检查、从伦理审查委员会(IRB)申请批准以及进行试验前研究等等。本书为决定在创建特定实验中最佳使用哪些要素以及在做出不可避免的权衡时需要考虑哪些因素提供了指南。本书的内容具有实用性和应用性。

《社会科学实验设计》只简要介绍了实验的统计方面，因为它们是密不可分的。然而，重点主要是让读者学会从一开始设计实验。本书不涉及统计的内部工作原理、公式以及如何手动计算或使用软件进行计算。了解统计学的深入知识并不是理解本书内容的必要条件。本书着重于帮助读者理解哪种统计方法适用于什么样的分析和测量水平，但内容非常基础。我建议除阅读本书之外，还参加传统的实验设计课程、统计学课程或阅读专注于统计学的其他书籍。

本书与其他书籍的第二个不同之处在于它面向广泛的社会科学研究人员。读者会发现，除了心理学之外，还包括政治学、商业学、经济学、信息科学、社会工作、教育学、社会学、卫生领域、广告学等学科的范例。我所从事的学科是新闻学和传播学领域。我教的第一门实验方法课程的学生来自新闻学、广告学、公共关系、广播、电视、电影、政治学和人际交流等不同领域。随着研究人员逐渐意识到理解因果关系对他们的学科的重要性，实验在社会科学中变得越来越普遍。随着易于使用的软件（如SPSS、G\*Power、Qualtrics）的引入以及对人类研究对象的更广泛接触（如Mechanical Turk），实验设计越来越普遍。本书直接针对这些研究人员，提供了来自他们自己领域的范例。它还探讨了特定学科面临的一些挑战，例如教育领域中对随机分组的反对意见，政治学中样本回应率的报告，以及使用观察或性能为基础的方法而不是结果为基础的方法。2

社会科学领域的许多研究人员通过与其他实验人员一对一合作来学习。本书旨在为那些通过这种方式学习的人提供补充材料，对于那些独立工作的人来说，它是一本指南，对于有组织课程的学生来说，它是一本教科书。我之所以写这本书，是因为我在德克萨斯大学教授实验方法课程时找不到完全符合我要求的教材。该课程是本书的基础。我将其设想为独立的实验设计方法课程的核心教材。对于在实验中占主导地位的广告学、政治学、健康科学等社会科学的一般研究方法课程，它也是一本很好的辅助教材。

本书是以研究生为目标读者编写的，尽管我在新闻界的十五年职业经历使我能够为广大读者撰写易于理解的语言。这本书的语言简单易懂，即使本科生也能理解。然而，对于完全没有经验的人来说并不合适；本书更适合受过教育的初学者，即以前进行过研究但没有进行过实验的人，或者参与过实验但没有自己进行过实验的人。它假设至少进行过一门关于科学方法的概述课程。那些具有一些研究经验但以前从未进行过实验的人不会觉得太具挑战性。对于那些通过导师而不是有组织课程进行学习的人来说，本书是作为参考和补充材料的，用以辅助那些出色的导师。它还可以帮助产生一些工具和策略的想法，这些想法导师可能没有常规地包含在他们的教学内容中。其他读者可能包括在其他研究方法方面有经验的学术人员，他们希望扩展他们探索因果关系的能力。此外，对于从事研究工作的专业人士，无论是为政治候选人、广告机构、教育机构还是其他行业工作，本书都将很有帮助。我希望它能激发那些没有独立实验课程的社会科学系创设此类课程。

各章包含了概念性和理论性问题以及实用的操作建议。它采用逐步方法，引导学生撰写假设和研究问题，设计刺激材料，并撰写方法部分。许多例子来自于我的个人经验，使我能够提供为何选择某些程序而不选择其他程序的背景故事，研究如何发展以及应对审稿人意见的建议。这些内容通常没有出现在发表的论文中。我还借鉴了其他社会科学领域进行的一些出色实验，尽管我无法提供关于选择的内部知识。

虽然各章的顺序大致按照设计实验的顺序进行展示，但读者可以根据需要选择阅读顺序。例如，在阅读第2章有关欺骗的内容后，有些读者可能希望直接跳到第11章，在该章节中讨论了与伦理审查委员会(IRB)相关的更广泛的伦理问题。为了便于无序阅读，书末附有术语表，其中列出了每章的关键术语，这些术语在文中以粗体标记并进行了定义。其他教学工具包括三种类型的方框特色：

**­­xv**

• “更多关于…”框展开了章节中某个问题的详细内容，以满足那些希望了解更多细节的读者的需求。

• “研究重点”框突出展示了特定实验在处理该章节主题方面的方式。

• “如何操作”框提供了执行章节中步骤并撰写论文的指导和示例。

此外，还有其他三个特色：

• “常见错误”是一个简短的项目符号列表，列出了初学者常犯的错误。

• 每章末尾的“检测理解”问题是一个简短的多项选择测验，以便您可以了解自己对材料的理解程度。

• “应用练习”是对更长的实践任务的说明，让您可以应用所学知识，包括逐步创建自己实验的研究提案。这个过程被分解成易于管理的部分，使任务不再显得那么艰巨。

每章还包含了有关主题更深入阅读的建议阅读列表。每个章节的一个显著特点是提供关于如何描述刺激和工具、与合著者合作、提交到期刊并回应审稿人常见问题的实用建议和示例。书中还提供了创建Twitter动态或Facebook帖子、在线随机选择器和进行功效分析的免费软件的链接。有些主题并不是实验特有的，但我努力针对这种方法的特定问题进行了深入探讨。例如，第3章解释了理论在实验中的作用，其优势在于测试和拓展理论。它还介绍了如何使用来自文献的经验证据来为被测试的假设和理论过程构建案例，并提供了类似绘制句子结构的假设书写的“公式”。

伦理在第2章中首次被引入，自然而然地紧接着叙述了斯坦利·米尔格拉姆的顺从权威实验和菲利普·辛巴多的监狱实验的故事，这两个实验引起了人们对实验科学中伦理问题的重视。这里的“更多关于…”框深入探讨了实验中比其他方法更具问题性的欺骗行为。伦理讨论贯穿于每一章节，例如，在第7章关于随机分组的讨论中，讨论了一些将其视为不道德的反对意见。该主题在第11章中以更详细的方式重新讨论，使用IRB作为框架考虑到其标准与伦理紧密相关。由于本书采用了逐步方法，因此IRB是在设计实验并研究人员必须申请批准之后才进行讨论的。然而，对伦理实验的重视并不是一个事后想法，而是在第11章中包含的，因为IRB是执行伦理行为的主要手段，所以这一章似乎是同时讨论两者最合理的章节。

关于实验的所知所学并不仅止于本教材；我鼓励通过阅读“建议阅读”部分中的一些优秀文章，阅读自己领域的实验以及与他人合作来继续学习。我希望您能喜欢并找到本书有用。

为了获得与本书相配套的额外教学和学习资源，请访问study.sagepub.com/coleman网站。

教师可以找到按章节划分的PowerPoint®幻灯片，以辅助讲课准备，还有一个包含多项选择和开放式考试题的Word®测试题库。学生可以在整本书中找到SAGE期刊文章的全文版本。要获取与本书配套的教学和学习资源，请访问网站

#### 致谢

For Review-No Commercial Use(2023)

我要感谢我的导师们，他们在我还是研究生时就传授给了我实验的热情——Esther Thorson和Glenn Leshner。感谢你们的指导、耐心和分享你们的激情。感谢你们告诉其他人（结果传到了我这里）：“她总是言出必行”，这有助于确保这一点的实现。感谢你们作为榜样，每次新的被试坐在椅子上时都会测量屏幕到椅子的距离，让我认识到细节的重要性。感谢你们总是抽出时间听取我关于新项目的介绍并给予建议。我还要感谢我的长期研究伙伴Denis Wu，我很幸运在我的第一个教职位置上遇到他，他一直与我并肩工作。他总是在我不擅长的领域表现出色，并在我们自我引导的实验之旅中一直是我的坚实支持。我要感谢所有与我一起进行实验的优秀研究生，包括但不限于Ben Wasike、Lesa Hatley-Major、Rebecca McEntee、Lewis Knight、Carolyn Yaschur、Trent Boulter、Viorela Dan、Angela Lee、Raluca Cozma、Avery Holton、Dani Kilgo、Siobhan Smith、Kate West、Joseph Yoo和Ji Won Kim等等。我还要感谢所有参加我实验设计课程的人，特别是来自其他部门的人，你们让我看到了你们领域的新理论和独特挑战，也允许我在撰写本书的过程中对你们进行实验。我从你们那里学到的知识与你们从我这里学到的一样多，甚至更多。

最后，没有所有审稿人的时间和专业知识，本书是无法完成的。他们温和地指出了缺陷和不足之处，引导我找到宝贵的资源，并耐心地帮助我使本书更加出色。我还要感谢SAGE的高级编辑Helen Salmon，她对这个提案表示了支持，并提出了许多重要的补充意见；感谢Jeremy Shermak为本书创建了网站；感谢SAGE的优秀团队，包括Chelsea Neve、Megan O'Heffernan、Eve Oettinger、Diane Wainwright以及研究方法、统计学和评估部门的其他成员，他们在本书的完成过程中给予了很好的支持。

SAGE非常感谢以下审稿人对本书提供的反馈意见：

Richard E. Adams，肯特州立大学；

Francis O. Adeola，新奥尔良大学；

Anna Bassi，北卡罗来纳大学；

Jacqueline Craven，德尔塔州立大学；

Sarah A. El Sayed，阿灵顿德克萨斯大学；

Janet Reid Hector，罗格斯大学；

Kyle J. Holody，滨海卡罗来纳大学；

Glenn Leshner，俄克拉荷马大学；

Ji Hoon Ryoo，弗吉尼亚大学；

Gerene K. Starratt，巴里大学；

Michael Teneyck，阿灵顿德克萨斯大学；

Geoffrey P. R. Wallace，华盛顿大学。

###### 注解

1. W. A. McCall，《教育实验方法》（纽约：麦克米伦，1923年），前言。

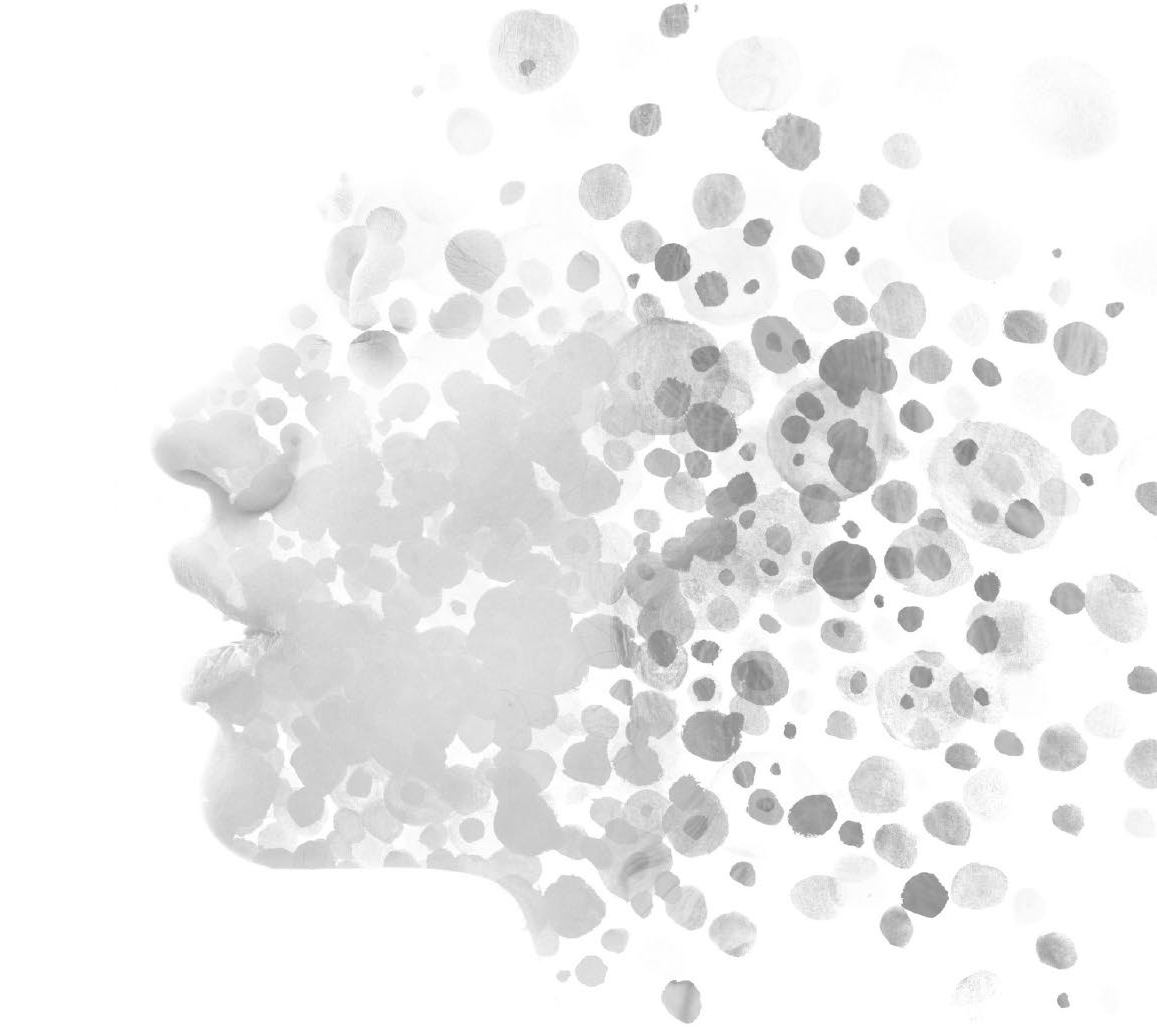
2. D. T. Campbell和J. C. Stanley，《研究的实验与准实验设计》（芝加哥：兰德·麦克纳利，1963年），1页。

## 关于作者

Renita Coleman在佛罗里达大学获得新闻学学士学位，并在密苏里大学获得新闻学硕士和博士学位。她的研究重点是伦理学、框架理论和议程设置，特别关注视觉传播。她研究了照片对伦理推理的影响、健康新闻中的框架和责任归因，以及新闻从业者的道德发展等课题。

她在学术期刊上发表了40多篇经过同行评审的文章，包括《传播学杂志》、《新闻与大众传播季刊》、《广播与电子媒体杂志》、《大众传媒伦理学杂志》、《新闻学》和《新闻学研究》。她合著了两本书：《选民决策中的形象与情感：情感议程》（2015年）和《道德媒体：记者如何思考伦理》（2005年）。在开始学术职业生涯之前，Coleman在报纸和杂志上担任记者长达十五年。她曾在北卡罗来纳州的《新闻和观察家报》、佛罗里达州的《先驱报》以及佛罗里达州的《奥兰多哨兵报》等新闻机构担任记者、编辑和设计师。

**xxi**



1

**发现因果关系**

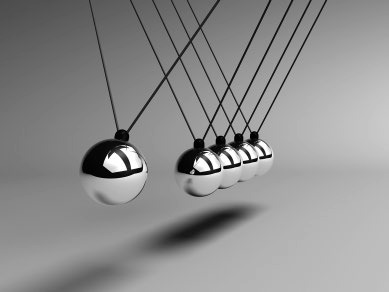
**生活是对因果关系不断探索的不断教诲。**

**——拉尔夫·瓦尔多·爱默生**

• 解释实验中的因果关系是如何运作的。 • 比较实验与其他方法的优势。 • 确定实验的三个基本标准。 • 描述变异、混杂因素、对照组和分组的要素。 • 为你自己的研究制定问题陈述，并回答“那又怎样”的问题。

这本书是关于实验的，其范围变化很大。实验是在精心控制的条件下进行的对某个假设或原理进行科学测试，以便确定或发现某些未知的东西。实验提供了关于事物之间关系的见解，其中一个事物的变化导致另一件事发生。正如拉尔夫·瓦尔多·爱默生的开头引言所示，我们在日常生活中经常进行非正式的实验，甚至不知道自己在做实验。例如，如果你改变食谱中某种成分的数量，口味会变得更好吗？如果你喝低热量啤酒，你会减肥吗？只要你一次只改变一个成分，或者不增加运动量或食用低热量的所有食物，你可能会认为是那个成分或者低热量啤酒导致了口味的差异或者你的减重。

**1**



#### 因果关系

iStock.com/terrymorris

这些日常实验的例子说明了因果关系的概念。在前面的例子中，口感更好的烤宽面条或减肥就是效果。而“原因”是解释为什么会发生这些事情的原因——烤宽面条中加入更多的大蒜使其口感更好，低热量啤酒帮助你减肥。我们在日常语言中使用“原因”的词，比如“死亡原因”或“事故原因”。在实验中，这个意思并没有不同，但你也会看到使用因果关系或因果性这些术语。

当然，在本书的目的上，我们对系统化的实验比对简单的日常实验更感兴趣。在医学中，这些被称为临床试验或随机临床试验（RCTs；有关RCTs的更多信息，请参见附录1F.1）。在市场研究和市场营销研究中，它们也被称为A/B测试。社会科学中的实验术语可能与自然科学中的实验略有不同，但目标是相同的——找出对特定问题最有效的治疗方法（或原因）。在医学中，问题是疾病或疾病，而效果是治愈或改善。在社会科学中，我们寻求解决的问题可以是促进品牌认知的电视广告，教授唐氏综合症学生的策略，或帮助会计师更加诚实的干预措施。

为了做到这一点，社会科学家使用实验作为“证明的基本语言”。通过展示在改变某个因素的同时保持其他一切不变时会发生什么，实验为我们提供了因果关系的证据。这样，我们更有把握相信发生变化的因素是导致我们观察到的结果或效果的原因。

实验是医学领域最常见的研究类型，但在社会科学领域，它们可能是最少被使用的。在传播学期刊中，约有12%的研究采用实验设计。在国际关系领域，这个比例可能只有4%。在特殊教育领域，实验被提升为基于证据的专业要求提高质量和严谨性的答案。在政治科学等领域，实验也在稳步增长。随着社会科学领域对这种方法的好处的认识逐渐加深，数据分析软件变得更易于使用，技术使受试者更便宜、更容易招募，实验的使用应该会增加。

### 更多关于... 盒子1.1

随机临床试验Trials

RCTs基本上与真实实验相同，受试者被随机分配（即R）到治疗组或对照组，并接受一个或多个实验程序或药物。它们也被称为随机对照试验，语言纯粹主义者在包含对照组的研究中使用"controlled"一词，其中受试者接受安慰剂或不接受治疗。当使用"clinical"一词时，可能存在对照组，也可能不存在。可以推测，"clinical"一词的使用是因为这些研究是在医学患者或其他临床环境中进行的。"trial"一词被使用是因为正在研究的治疗方法或药物正在进行试验，也就是说，它尚未获得广泛使用的批准，研究将确定其是否安全有效。1

RCTs被认为是医学研究的金标准，就像在社会科学中真实实验或实验室实验一样。

RCT的历史可以追溯到公元前600年，当时犹大的但以理比较了巴比伦王室的饮食和素食。2其他人将其归功于詹姆斯·林德，他在1747年进行了有关坏血病的实验，这在第二章中有描述。3通常将第一次现代RCT识别为1948年关于链霉菌对结核病的影响的测试。4该研究首先解释了有关结核病临床试验缺乏充分控制导致有关黄金治疗的“夸大宣称”的情况。由于规划这样一项研究的困难性质，该研究提供了“完整的描述”。

因此，了解如何进行良好的实验显得更加重要。许多文章已经讨论了实验在学术界和专业领域的增加应用，例如政治学11和信息系统12。还有一些文章讨论了在实验方法方面创建有组织的课程的日益重要性13。有些人甚至认为实验是他们学科中最重要的方法14。实验在以循证实践为基础的领域，如社会工作15和教育中非常核心，其中随机对照试验是金标准16。

#### 与其他方法相比的实验

实验方法的主要好处在于它提供了一个强大的工具来发现因果关系。许多其他研究方法，比如调查，可以确定相关性或者一起变化并且不太可能是偶然发生的关系。然而，这并不一定意味着因果关系。相比之下，实验可以揭示改变一件事物如何导致另一件事物的变化。有意识地和系统地改变某个因素使我们能够观察到潜在的因果因素。通过在相关性研究之后进行实验，可以更有把握地确定这些研究中发现的关系是否真的具有因果关系。使用不同方法对同一现象进行研究的三角测量在科学研究中非常重要，因为它有助于增加我们对使用一种方法观察到的现象能够通过另一种方法得到重复的信心。目前令人担忧的是过于依赖调查和观察方法来显示相关性，但没有进行实验跟进来找到因果关系的证据。

从相关性推断因果关系的问题最著名的例子之一是很久以前的说法，即鹳鸟带来婴儿。上世纪30年代在哥本哈根进行的一项研究记录下了，鹳鸟数量较多的年份也出生了更多的婴儿，相关性达到了0.85的高水平。然而，仅仅因为这两个变量高度相关，并不意味着一个导致了另一个。相反，还存在其他可能的解释或者其他可能的原因没有被研究到。那是在第一次世界大战后，所有那些长时间离家的士兵回到家中导致了更多的婴儿出生。此外，人们从农村迁往城市找工作，因此更多的人生婴儿意味着更多的婴儿出生。随着人口增加，建造了更多的房屋，这为鹳鸟筑巢提供了更多的地方，导致鹳鸟数量增加。这个循环持续下去。

.

iStock.com/zuzulicea

#### 实验的基本标准

这个例子说明了所有实验必须具备的三个基本特征：即原因必须在效应之前发生，比如在这个例子中，鹳鸟确实在婴儿之前出现，但是原因必须与效应相关，而这并非如此。没有逻辑上或理论上的原因，也没有经验证据表明鹳鸟与婴儿有关。这也是为什么实验者不仅仅测试主观猜测，而是寻找理论、逻辑或已有证据来验证两者之间的关系。这有助于确保原因实际上与效应相关。在鹳鸟和婴儿的例子中，实验的第三个特征也缺失了，即对于效应，不能有其他合理的替代解释。一个真正的实验必须同时具备这三个条件：原因必须在效应之前发生，与之相关，并且不能有其他合理的替代解释。只有满足了这三个条件，实验才能为我们提供一种更有力的方式来获取更多的信息。确信某个因素导致或引起了另一个因素，而不仅仅是存在一个关联，这种关联很可能不是偶然发生的。

尽管能够测试因果关系是一个强大的工具，但它仍然只是一种描述，它并不能解释为什么某件事会发生。例如，在社会工作领域，许多研究已经确定了成功减少青少年犯罪的计划，但很少有人对它们的有效性进行了研究。Hay等人注意到这个空白，并设计了一个实验来发现解释参与计划如何降低犯罪率的机制（有关该研究的更多信息，请参阅研究亮点1.2）。在我的研究中，我假设观看照片会导致新闻记者在做出新闻决策时使用更高质量的伦理推理。我找到了我寻找的效应，但那只是一种描述，而不是一种解释。仅凭这一发现并不能说明为什么会发生这种效应。为了提供因果解释，实验需要建立解释这些效应的潜在机制。如果照片确实能提高伦理推理能力，了解为什么很重要。这样的中介和调节因素将在另一章节中详细讨论。简而言之，实验需要建立潜在的机制来解释这些效应。



### 研究亮点 1.2

### 发现效应并解释原因

coleman

**这项研究是一个很好的例子，不仅找到了效应，还解释了造成这些效应的原因。作者首先再次确认了特定计划的治疗确实降低了青少年犯罪的风险。然后，他们研究了介入变量，即参与计划和降低青少年犯罪之间的因素。他们总共考察了11个被认为是计划有效原因的风险因素，研究人员称之为因果机制。他们表示：“简而言之，如果我们缺乏对计划如何减少犯罪的具体机制的了解，那么在增强其优势、在其他地方复制它，并用其来制定公共政策方面的努力必然会受到阻碍”（第248页）。**

**在可能解释计划有效原因的11个潜在变量中，研究人员只发现了一个显著的变量：“与从事不良行为并对其施加压力的同伴关联减少”（第263页）。换句话说，与不良团体交往。他们解释了这个发现的重要性，涉及到这些计划在追求其他十个变量上所耗费的时间和金钱，而事实证明这些变量并没有实际产生差异。研究人员以较为友好的措辞表达了这一观点，他们说：“我们的分析——这是一个罕见的考虑到介入变量的测试——表明这些计划所针对的许多风险因素可能对计划服务没有反应”（第264页）。**

在这里可以说，良好的实验应该包括解释为什么某种效应发生的机制。

在对实验相对于其他方法的能力进行基本讨论之后，接下来我们将转向实验的一些具体要素。所有这些要素将在后面的章节中详细阐述，但在这里引入它们是为了提供对实验的基本基础理解。

#### 实验的要素

变量变化

变化是进行良好实验的首要关键。显然，如果没有变化，就没有可研究的内容。某些东西必须发生改变。在实验中，变化通过操作自变量（IV）来实现，即研究人员认为会引起变化的因素。这也被称为操纵、处理或干预。研究人员应该仔细控制这些变化。例如，许多研究使用专业演员来扮演政治候选人，而不是真实的政治人物，以避免可能对被试者回应产生偏见。在一项研究中，演员在不同的采访中展示了积极的非语言行为、消极的行为和中立的肢体语言。研究作者详细研究了这些行为的特征，例如，交叉双臂是消极的，向前倾身并目视采访者是积极的，并确保演员表现出这些行为。仅仅要求演员积极或消极地表现是不够的，研究人员需要确切知道这种行为应该是什么样的，并确保其正确表达。

For Review-No Commercial Use(2023)

实现良好变化的另一个关键是一次只能有一件事情发生变化；否则，就无法确定几个变化中的哪一个导致了结果。实际上，许多实验确实同时变化了多个因素（稍后详细介绍），但关键在于它们不允许这些因素共变，也就是说，它们不能同时变化。例如，有研究者注意到，显示负面政治广告的研究比积极广告更有影响力，并且广告的语调与其中的信息量共变。负面广告中的信息量始终比积极广告多；因此，更强的效应可能是由于更多的信息，而不一定是其负面语调。这为这些其他研究中发现的效应提供了一个合理的替代解释。为了确定是否属实，研究者进行了一项研究，将信息量保持不变，即负面和积极广告中的信息量相同。他发现，语调根本没有引起效应，而是信息量；

负面广告通常比积极广告包含更多信息，这就是其他研究所记录的更高竞选知识、兴趣和投票率的原因。

实际上，因果关系很少是单变量的，或仅由一个变量引起的。社会科学研究人员很少指望一件事情单独成为另一件事情的直接原因。通常有很多因素共同起作用。在实验中，为了确保这些因素不对效应负责，研究者会将除了被有意变化的变量之外的其他因素保持恒定。

混淆变量

可能提供合理替代解释的因素被称为混淆变量（confounds）。重要的是，实验中不应该存在混淆变量，即任何可能损害实验准确性的因素。成功的实验关键在于控制可能导致结果的外部影响因素。例如，一位研究人员研究报纸中位于折叠线以上或以下的照片对人们的影响，使用了一张位于折叠线以上展示近距离人物的纵向照片，以及一张位于折叠线以下展示远处人群的横向照片。他只使用了折叠线以下的横向远处照片和折叠线以上的纵向近距离照片。问题在于，导致他发现的效应的可能是照片中的近距离或远处人物的特征，而这些不是他有意变化的变量。这些照片在许多层面上发生了变化，不仅仅是他感兴趣的报纸位置。这构成了一个混淆变量，即一个未被控制的合理替代解释。

为了避免混淆变量，在一项旨在确定照片中人物种族如何影响记者伦理判断的研究中，研究人员使用完全相同的照片，通过数字修改人物的肤色、发型和面部特征来操纵其种族。这样，除了种族之外，其他一切都保持一致。背景相同，离摄像机的距离相同，人物的吸引力和其他方面也相同。这样避免了任何混淆变量。研究人员知道，例如，近距离拍摄会让人们对照片中的人物感到更舒适，而远距离拍摄则不会产生这种效果；有吸引力的人在各种特征上会得到更好的评价，包括政治候选人的可信度和当选能力。因此，这项研究需要确保除了人物种族之外，照片在其他方面完全相同。

在实验中，尽可能控制混淆变量是有帮助的。例如，研究人员可能会在每次新的实验开始之前测量受试者坐在电视机旁的距离。所有受试者应该与电视机的距离完全相同，因为即使稍微靠近一点，也可能产生影响。

研究人员可以通过常识和阅读其他研究来发现潜在的混淆变量。文献中有大量证据表明，例如，政党和意识形态会影响许多事情。因此，许多实验通过创建不暗示或明示虚构候选人政党的刺激来控制这些潜在混淆变量。这代表了实验控制，即精心管理参与的变量。避免潜在混淆的另一个例子是在选举期间的实验中，不使用当时新闻中的真实问题。当研究人员无法控制自变量和因变量之间潜在虚假关系时，另一种方法是测量并对其进行统计控制，即在检查操纵效应之前排除它。测量人们的政党和意识形态，然后将其作为协变量进行统计控制，是一种统计控制的形式。

了解领域内的文献是重要的，以发现需要控制的因素。例如，在道德发展的研究中，研究人员发现保守主义或自由主义与道德判断测试并观察。假设您正在研究一个您非常了解的现象。如果您是专家，应该对影响您研究的现象的各种因素有一些想法。这是知识生成和文献发展的方式之一。

对照组

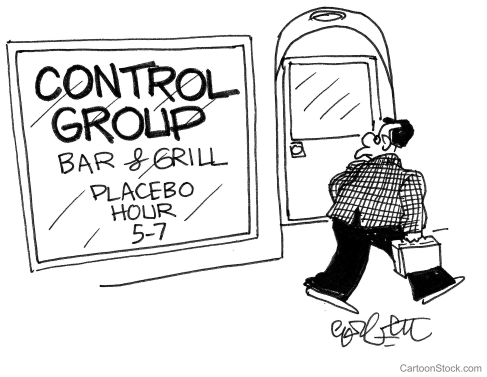
实验的另一个重要要求不仅是了解接受治疗的人会发生什么，还要了解如果他们没有接受治疗会发生什么。有一个不接受治疗的人群，即对照组，可以推断如果没有治疗，结果将会如何。效应是接受治疗的人群与未接受治疗的人群之间的差异。这样我们就能够分离和测试一次只有一个变量的效应，并且更加确定这种效应是由该变量而不是其他因素引起的。接受治疗或干预的人群被称为治疗组、实验组或干预组。没有接受治疗的人群则是对照组。对照组是没有接受干预的人群，被用作与“正常”、“中性”或“无干预”情况进行比较的参照组。

在医学研究中，对照组通常会接受安慰剂，它看起来像是一颗药丸、注射、疗法或其他治疗方式，但实际上只是一颗糖丸、盐水注射液或其他让人感觉有接受治疗的物质。（有关安慰剂的更多信息，请参阅“更多关于盒子1.3的内容”。）由于暗示的力量，仅仅对对照组的人什么都不做是不够的；他们必须感知到他们接受了某种治疗。对于社会科学家来说，这有时可能会带来问题。例如，在测试不同类型的营销信息时，什么是对照组？当然不发送信息，但那么受试者做什么呢？不能让他们只是进来回答问卷而没有暴露于某种干预，否则他们会认为研究过于人为。对照组将在接下来的章节中进一步讨论；这里的讨论重点是实验需要某种方式来比较接受治疗的人群和未接受治疗的人群之间的差异。

分配

这次讨论的最后一个话题，但可能对于实验来说最重要的是人们如何被分配到实验中的不同干预条件。这个话题非常重要，以至于它在本书中有自己的一章（第7章）。简要地说，黄金标准是随机分配，这是一种将被试者（注：现代研究更常使用"participants"或"individuals"一词来强调对人类研究对象的道德对待）以一种方式分配到各个组别中，以便个体差异均匀分布在不同的组别中。这一特点非常重要，以确保人们的个体特征不会成为研究的混淆因素。例如，男性和女性应该在各个组别中均匀分布，年轻人和老年人也应该如此。本书专门为将被试者分配到实验条件的问题撰写了一整章，包括讨论在无法按照这种方式分配人员时应该采取的措施。

### 更多关于... 盒子1.3



### 安慰剂

在医学研究中，安慰剂的使用比简单地注射生理盐水、服用糖丸或模拟手术等方式对照组进行假装治疗要复杂得多。事实上，这些无效但无害的治疗实际上已被证明具有效果。被称为安慰剂效应，接受安慰剂的人在被假定接受治疗的任何疾病中报告症状改善。

安慰剂一词源自拉丁语“我将使人满意”。33第一次使用安慰剂进行对照试验可以追溯到1801年。34关于安慰剂的欺骗性、虚假性和伦理问题的讨论始于20世纪中叶，当时许多执业医生会在真实药物的幌子下向患者施用安慰剂。35对于在研究中使用安慰剂的伦理性的问题至今仍存在争议。36

关于安慰剂效应的第一份报告发现，接受安慰剂的患者中有五分之四报告症状缓解；当他们接受实际治疗时结果是相同的。37该报告提到了希望、信念和想象力的力量。自那以后，关于安慰剂效应的研究一直持续显示，患者在接受无效的安慰剂后程度不同地改善，包括与有效药物相同甚至更好的效果。38对15项研究的综述表明，平均而言，安慰剂在35%的情况下与有效治疗具有相同的效果。39这个综述的缺点已被指出，但研究继续表明安慰剂在约三分之一的情况下“有效”。40

1961年，Beecher进行了另一项综述，发现37%的患者存在安慰剂效应。41安慰剂效应一词被定义为想象力的力量导致无效物质的积极效果，42或者在一个公正的实验中，安慰剂治疗组与未接受治疗的对照组之间结果的差异。43

在社会科学研究中，安慰剂类似于接受一种看起来像是治疗或干预的对照组44，例如阅读关于一部新电影的故事而不是研究的真正焦点——连续和主题相关的犯罪故事。

安慰剂是给予被试者的无效或虚假治疗；安慰剂效应是被试者对它的反应。这与第2章中的霍桑效应不同，后者描述了由于被观察而导致的被试者表现的变化。在安慰剂效应中，被试者的改变是因为他们对治疗的期望、信念或希望，而不是治疗本身。45为了有效，一种治疗必须通过安慰剂测试——也就是说，该治疗在结果上必须显著优于接受安慰剂的对照组。社会科学的研究也发现了安慰剂效应，例如与承诺提供更好运动表现的产品46、影响金融决策的因素47、饮品标签与醉酒感知48以及知名品牌49。安慰剂也存在于日常生活中，例如路口的按钮让行人产生一种控制感，但实际上并不真正影响信号灯50。

#### 开始您自己的研究

现在我们对实验相对于其他方法的基本理解有了一些了解，本章将转向启动您自己的实验的第一个实际步骤：撰写清晰简明的研究问题陈述。在各种研究中，清晰简明的写作都很重要，但根据我的经验，对社会科学研究人员来说，特别是对于那些对实验方法可能不太熟悉的人来说，实验似乎更难以理解。因此，本书的一个关键重点是帮助读者用简单的语言和任何具有中等社会科学知识的人都能理解的术语来撰写实验。这可能比您想象的要困难。我们从研究问题的陈述开始。这个问题的范围比具体的假设更广泛，尽管它们是相关的。编写假设将在后面的章节（第3章）中详细介绍。开始一个新实验，您必须提出一个要测试的想法。在思考时，将其以因果关系的方式考虑会有所帮助，即测试一件事是否导致另一件事发生。例如，这里有三个（相对简化的）研究想法：

ii本教科书使用术语"subjects"来指代实验中的研究对象。美国心理学协会（APA）最新（第6版）的《出版手册》指出，"subjects"和"participants"两个术语都是合适的，并指出"subjects"一词已经使用了数百年。对于"subjects"一词的反对意见始于1994年APA手册的第四版，当时更倾向于使用"participants"。当当前的第6版问世时，"subjects"和"participants"并列为合适的术语。手册第73页的条目指出：“事实上，100多年来，“subjects”一词一直在实验心理学中作为描述样本的一个通用起点使用，并且其使用是合适的。”因为"subjects"和"participants"同样合适，本教科书使用"subjects"以保持与实验设计语言中的其他术语的一致性，包括组间设计（between-subjects designs）、组内设计（within-subjects designs）以及伦理审查委员会使用的人体研究对象（human subjects）。有关这个主题的更多信息，请阅读Roddy Roediger在2004年4月出版的APS Observer杂志上的文章“What Should They Be Called”，可在http://journal.sjdm.org/roediger.html 上获取。

1. 看照片是否能提高伦理推理能力？

2. 声音音调是否会影响广播播音员的可信度？

3. 女性政治家的身高是否会影响选民对其资格评估和投票可能性？

在第一项研究中，治疗组接受展示照片的条件，而对照组则没有照片。在第二项研究中，广播播音员的声音在一个治疗条件中被调高，而在另一个治疗条件中被调低，对照组的声音保持正常。在第三项研究中，矮个子女性和高个子女性的照片被分为两个治疗组，而平均身高女性则作为对照组。当然，这三个实验的实际情况要比这个简单的“A是否导致B”的方法更为复杂，但这种简单的方法是所有研究的起点。

注意：本回答基于您提供的信息进行翻译，实际研究的设计和操作可能会更加复杂和详细。



撰写问题陈述

下一步是撰写一个明确而聚焦的研究问题陈述。一个好的开头方式是使用“本研究的目的是...”或类似的句子。以下是一些来自实验研究的好的、清晰和聚焦的问题陈述的例子：

Mine Cetinkaya-Rundel

For Review-No Commercial Use(2023)

• “本文研究了微博工具Twitter在媒体使用中对所涉问题的认知和信息的可信度的影响。”51

• “我们实验性地研究了金钱是否能购买政治影响力这个常见的智慧。”52

• “我们评估了沟通环境（即面对面与在线聊天室讨论）对个体表达意见意愿的影响程度。”53

• “首先的问题是，提供关于价格和市场福利特性的一般信息是否会改变对不可接受交易的态度。”54

• “. . . 本研究的目标是证明道德信念和政治道德判断受到伤害关联和道德情感的因果影响。”55

注意：本回答基于您提供的信息进行翻译，实际研究的问题陈述可能会更具体和详细。

所有这些例子有三个共同点：它们都指出了干预、操作或因果关系是什么，以及结果或效果是什么，通常是按照这个顺序。并且它们都包含一个动词。在第一个例子中，Twitter是干预或因果关系，效果或结果是对问题的认知和可信度。在下一个例子中，金钱是原因，政治影响是结果。在最后一个例子中，顺序被颠倒，效果——道德信念和判断——排在第一位，原因——伤害关联和道德情感——排在最后。您也可以将其识别为自变量（原因、干预或操作）和因变量（效果或结果）。共同点的第三个是每个例子中都有一个动词，用来描述原因对结果的作用，或者一个事物对另一个事物的影响。在这些例子中，动词分别是“影响”、“购买”、“修改”和“因果影响”。在实验中，建议避免使用不太精确的词语，如“探索”、“理解”和“研究”，而是使用更具体、因果关系的语言。其他可以使用的好词包括“不同”、“改善”、“提高”和明确的动词。例如，可以这样开始一个实验的研究问题陈述：“本研究的目的是测试照片是否改善了道德推理”，或者“本研究的目的是探讨广播员声音的音调是否会影响对可信度的评估”。要为实验撰写一个明确而聚焦的研究问题陈述，首先确定原因（或干预、操作、独立变量），然后说出预期的影响（不同、影响、修改、引起——动词）以及某种结果（效果或因变量）。在这里，我提供了一个填空模板，用于撰写研究问题陈述：

“本研究的目的是探究（某个因素，插入原因、干预、操作、处理或独立变量）如何（做某事，插入动词——不同、影响、修改、引起等）（某个结果，插入效果、结果、因变量）。”

如果这些例子中的某些术语，如伤害关联，不熟悉，那么检查您所在领域的一些实验，寻找其他更常见的例子可能会有所帮助。还要注意到所有这些例子都是清晰、具体和聚焦的，描述了实验的目的和问题陈述。撰写这样一个明确、聚焦且易于理解的陈述，用一句话描述实验，并不容易。很可能需要多次写作、修改和编辑。请熟悉该学科的人阅读并确认是否能理解。随着项目的发展，您可能需要重写这个陈述，甚至可能需要多次。

许多实验有多个目的，因此可以编写两到三个问题陈述。如果是这种情况，请将它们放在一起，并使用诸如“本研究还旨在……”之类的短语进行衔接。例如：“本研究测试照片是否改善道德推理。如果发现这种效果，本研究还旨在确定这种改善的因果机制。”这样，主要目的与其他目的都在同一个地方呈现，而不是分散在整篇论文中。读者会感激将研究意图的所有内容都列在一起，而不是阅读一个逐渐演变的研究目的，像是等待一个谋杀悬疑小说的情节逐渐展开。

在满意研究问题陈述后，作者需要定期提醒读者研究的使命，因此请确保问题陈述保持一致。我在审阅论文时经常看到一个问题，随着论文的进展，实验的问题陈述会发生变化。

最后，一个好的做法是将问题陈述写在便利贴上，粘贴在您的计算机上，以便您在工作时可以看到。这将有助于保持专注和一致。撰写这样一个明确、聚焦且易于理解的陈述，用一句话描述实验，并不容易。很可能需要多次写作、修改和编辑。请熟悉该学科的人阅读并确认是否能理解。随着项目的发展，您可能需要重写这个陈述，甚至可能需要多次。

许多实验有多个目的，因此可以编写两到三个问题陈述。如果是这种情况，请将它们放在一起，并使用诸如“本研究还旨在……”之类的短语进行衔接。例如：“本研究测试照片是否改善道德推理。如果发现这种效果，本研究还旨在确定这种改善的因果机制。”这样，主要目的与其他目的都在同一个地方呈现，而不是分散在整篇论文中。读者会感激将研究意图的所有内容都列在一起，而不是阅读一个逐渐演变的研究目的，像是等待一个谋杀悬疑小说的情节逐渐展开。

在满意研究问题陈述后，作者需要定期提醒读者研究的使命，因此请确保问题陈述保持一致。我在审阅论文时经常看到一个问题，随着论文的进展，实验的问题陈述会发生变化。

最后，一个好的做法是将问题陈述写在便利贴上，粘贴在您的计算机上，以便您在工作时可以看到。这将有助于保持专注和一致。

回答“那又怎样”的问题

在制定了这个明确而专注的陈述之后，下一步是阐述为什么这项研究很重要。这通常被称为回答“那又怎样”的问题。有些期刊甚至有一个特别突出的部分专门用来阐述这一点——例如，参见《美国国家科学院院刊》的“意义”部分。这是研究人员常常认为他们的研究的重要性是显而易见的，不需要指出的另一个领域；事实上，这是一项最关键的任务之一。对于这个任务，我告诉学生他们需要陈述显而易见的事实。对于你来说可能非常明显，但对其他读者来说并不一定如此。为了做到这一点，可以从三个不同的层面考虑你的研究为什么重要：（1）对其他学术界人士的重要性，（2）对该领域的专业人士的重要性，（3）对社会或一般人群的重要性。对于第一个层面，可以指出某个知识空白或研究克服的障碍。

此外，研究还可以在某种程度上对理论做出贡献，或者帮助揭示某些现象的机制或原因。有一种原因单独来看是不可接受的，那就是因为以前从未进行过这样的研究。这只是一个起点，你应该进一步阐明为什么除了以前从未进行过这种研究之外，它的重要性。否则，也许以前从未进行过这项研究是有充分理由的。

经常看到研究人员强调对其他学者的重要性，但忽视了其他两个群体：专业人士和公众。对于任何类型的研究，能够清晰地阐明你的研究结果对于学术界以外的领域的意义至关重要，这对于有实际影响的研究至关重要。R. Barker Bausell的书《Conducting Meaningful Experiments》正是基于这个前提。在考虑“那又怎样”的问题的第二步中，请问你的研究对你所在学科的专业人士是否感兴趣？它是否能帮助会计师、竞选经理、教师、公关专业人士或其他人更好地完成工作？研究结果是否能够让专业人士更深入地了解自己潜意识决策的过程？它是否能帮助他们克服一些障碍或提供他们改变实践方式所需的证据？也许它能够告诉他们哪些努力能够达到他们期望的结果，哪些努力则不能，无论是更积极参与的公民还是更多的客户。通过进行这种有针对性的努力，可以在学者和服务对象之间架起沟通的桥梁。

最后，能够说明为什么整个社会都会受益将产生更有意义的科学研究。公众往往认为学术研究人员是“象牙塔中的书呆子”，他们写的东西没有现实基础，只是为了发表论文。甚至有一些奖项因最浪费资源的研究而成为新闻。如果研究能够更好地阐明为什么看似愚蠢或显而易见的发现对于学术同行以外的其他人来说很重要，就可以避免自己的研究遭受这种嘲笑。这种关注对于推进研究事业或增加研究资金没有任何帮助，而进行对普通人有意义的研究则是有意义的。所以要回答“那又怎样”的问题：它是否会帮助任何人或改善某些社会问题？并不是每项研究都能达到世界和平的目标，但它可能有助于减少种族歧视、改变一种道德上令人反感的做法，或者为决策者提供制定法律所需的信息。并不是每项研究都能同时阐述这三个层面，但通过深思熟虑对这些不同受众的益处，实验研究人员可以设计出真正有意义的研究。有关回答“那又怎样”的问题的陈述示例，请参见《如何做》1.4节。

**HOW TO DO IT 1.4**

Examples of Answers to the “So What” Question

**从：Neil, Nicole和Emily A. Jones. 2015. “Studying Treatment Intensity: Lessons from Two Preliminary Studies.” Journal of Behavioral Education 24: 51–73.**

**“只有最近的一些研究关注了治疗强度，并且对于特定疾病的治疗强度的研究更为有限。也许特定病因和相关特征会影响干预强度的效果。许多患有唐氏综合症的儿童表现出差的任务持续性和不一致的动机导向...对于一些患有唐氏综合症的学习者来说，可能存在适度强度的最佳水平，在此水平之后，学习者参与更多的逃避行为和获得速度的增益递减。”59**

**从：Coleman, Renita. 2011. “Color Blind: Race and the Ethical Reasoning of African Americans on Journalism Dilemmas.” Journalism and Mass Communication Quarterly 88 (2) (夏季): 337–351.**

**“这项研究具有价值，因为它提供了评估解决刻板媒体形象问题所提出的一个解决方案的重要信息 - 雇佣和提升更多的少数民族记者。全国各地的新闻编辑部主要由白人组成；根据论点，加入更多的少数民族观点应该导致对少数群体更平等的报道。因此，重要的是要考察少数民族记者在认知加工中是否确实表现出对少数群体更宽容的态度。迄今为止，没有实证证据表明黑人记者对新闻中的黑人有更积极的看法...这项研究还填补了这方面的空白，探讨了种族如何影响黑人在涉及黑人和白人的新闻报道中的道德推理。”60**

**从：Aday, Sean. 2006. “The Framesetting Effects of News: An Experimental Test of Advocacy Versus Objectivist Frames.” Journalism and Mass Communication Quarterly 83 (4) (冬季): 767–784. “具体而言，我们对于为什么在某些情况下发现效果而在其他情况下却没有发现效果了解还不够，而且对于探索这种效果的认知基础的研究还很少，这将使我们能够建立一个关于一些属性何时以及为何具有二级效应的理论。”61**

**从：Elias, Julio J., Nicola Lacetera和Mario Macis. 2015. “Markets and Morals: An Experimental Survey Study.” PLoS ONE 10 (6) (6月1日): 1–13. Public Library of Science.**

**“禁止其中一些交易是有代价的。例如，寿险合同曾经被视为违背上帝的赌博而被禁止，但现在它们为数百万人创造价值，并被视为一种‘制度化的利他主义’。同样，虽然有关其效率的论证早就提出，但在美国长期以来一直拒绝实行全志愿付费军队的想法。禁止对捐献器官的人进行支付导致器官需求和供应之间的差距日益扩大。禁止某些交易也可能导致非法市场的形成，这反过来又带来了进一步的成本，如暴力。”**





**从：Grober, Jens, Ernesto Reuben和Agnieszka Tymula. 2013. “Political Quid Pro Quo Agreements: An Experimental Study.” American Journal of Political Science 57 (3) (7月): 582–597.**

**"有充分的理由怀疑政治中的金钱不受欢迎的普遍观念背后有一定的真实性。首先，尽管被禁止，政治上的交换可以在公开观察不到的渠道之外发生。其次，对于那些经济实力强大的特殊利益集团，其中大多数是大型企业公司，以增加利润的形式给予支持更有可能是一种比政治参与更合理的解释。此外，对这些利益集团的回报，如特定的税收减免、补贴和法规，可以轻易地掩盖为经济上的必要性，因此很难量化。第三，主要候选人之间的勾结也可能采取关于某个特定政治问题的共同观点的协议形式......最后，即使公众对政治中金钱的影响被夸大了，这种信念仍可能影响公众的政治信任和行为。"63**

**常见错误**





在论文中没有清晰地陈述研究的目的，并且没有在整篇论文中保持一致。

没有将研究的所有内容逐步呈现出来，而是在研究进行过程中逐渐揭示研究目标。

在论文中没有在开头和第3页之前清楚地陈述研究目的。

未能说明为什么该研究对理论、其他研究人员、职业和普通人重要。"因为从未有过类似研究"本身并不是一个研究重要性的理由。

1. 实验需要表明因果关系先于效果，但不一定与效果相关。如果两者之间存在统计学上的显著关系，那就足够了。

a. 正确

b. 错误

2. 一位研究人员研究了吸引力对学生对教师的喜好程度的影响。吸引力较高的教师年龄为25岁，吸引力较低的教师年龄为45岁。这里的问题是：实验需要展示因果关系先于效果，但不必证明二者必然相关。如果两个事物之间存在统计上的显著关系，那就是最重要的

**检测您的知识**

**18** Designing Experiments for the Social Sciences

a. 年龄与吸引力存在混淆变量

b. 很难定义吸引力

c. 因果关系不在因果前

d. 研究人员没有控制政治意识形态

3. 你将15名员工分配到为期一天的压力管理研讨会，另外15名员工分配到为期一周的

研讨会。一个月结束时，你测量每位员工的感知压力水平。这项研究中的处理或操纵

是什么？

a. 员工的压力水平有多高

b. 研讨会的持续时间有多长

c. 如何选择这30名员工

d. 压力管理导师的质量

4. 一项研究在学生参加考试前测量他们的兴奋水平。结果发现，随着兴奋水平的增加，

表现下降。这一发现表明：

a. 因果关系

b. 相关关系

c. 一个合理的替代解释

d. 混淆变量

5. 以下哪项不是实验的三个基本标准之一？

a. 因果必须先于效果发生

b. 效果不太可能是偶然发生的

c. 因果必须与效果相关

d. 对于效果没有合理的替代解释

6. 变异是通过以下哪种方式实现的？

a. 保持所有条件不变

b. 使用人口统计作为协变量

c. 系统地改变某些事物

d. 设置一个对照组

7. 可能提供合理替代解释的事物被称为：

a. 协变量

b. 混淆变量

c. 独立变量

d. 因果机制

Chapter 1 ■ Discovering Cause and Effect **19**

8. 在实验中，对照组的作用是：

a. 让我们了解接受处理的人会发生什么情况

b. 让我们能够推广到更多的人

c. 让我们知道如果受试者没有接受处理会发生什么情况

d. 让我们能够以特定程度的确定性说某种效果发生了

9. 在实验中，“分配”是指：

a. 受试者必须完成的任务

b. 论文中作者顺序的计算方式

c. 研究人员确保受试者相信实验是真实的方式

d. 受试者被分配到不同的干预或对照组的方式

10. 在实验中，分配受试者的黄金标准是：

a. 代表性地分配

b. 故意地分配

c. 随机地分配

1. In an experiment, “assignment” is:
   1. The task that subjects must complete
   2. How authorship order is calculated for the paper
   3. The way researchers ensure subjects believe the experiment is real
   4. How subjects are put in the different interventions or control group
2. The gold standard in experiments is to assign subjects:
   1. Representatively
   2. Purposively
   3. Randomly

d. 随意地

答案:

1. b

2. a

3. b

4. b

5. b

6. c

7. b

8. c

9. d

10. c

1. 使用scholar.google.com或您学校图书馆的数据库，在您的学科中查找使用实验设计的研究。在搜索词中包含实验、实验设计或对照实验。阅读三个您最感兴趣的实验，并查找涵盖的概念。具体来说，确定治疗组或治疗组群。是否有对照组？如果有，用什么来表示“无治疗”？确定问题陈述和对“那又怎么样”的问题的答案。

2. 构思三个不同的研究，您希望通过实验来进行这些研究。也就是说，应该通过操纵或改变某些因素，以观察其对某个结果的影响。清晰而专注地陈述问题。解释为什么这对学术界、职业界和整个世界都很重要（“那又怎么样”的问题）。每个问题使用250个字来表述。这些问题不应该是简单的复制，而应该是新的想法，或者它们可以是对您复制的研究进行重要扩展的复制研究。

**应用练习**

**20** Designing Experiments for the Social Sciences

The introduction and Chapters 1 and 2 of Bausell, R. B. 1994. *Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist*. Thousand Oaks, CA: Sage.

Chapter 1 in Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth Cengage Learning.

Thorson, Esther, Robert H. Wicks, and Glenn Leshner. 2012. “Experimental Methodology in Journalism and Mass Communication Research.” *Journalism and Mass Communication Quarterly* 89 (1): 112–124.

**建议阅读**

###### 注释

1. Markus MacGill，《医学研究中的随机对照试验是什么？》，Medical News Today，http://www.medicalnewstoday.com/articles/280574.php。访问日期：2017年2月10日。

2. Harald O. Stolberg、Geoffrey Norman和Isabelle Trop，《放射科医生的临床研究基础：随机对照试验》，《美国放射学杂志》，第183卷，第6期（2004年）：1539-1544。

3. Marcia L. Meldrum，《随机对照试验的简史：从橙子和柠檬到黄金标准》，《北美血液/肿瘤学临床杂志》，第14卷，第4期（2000年）：745-760。

4. S. Shikata等，《消化外科随机对照试验与观察研究效果的比较》，《外科学年鉴》，第244卷，第5期（2006年）：668-676；Stolberg、Norman和Trop，《放射科医生的临床研究基础》。

5. 结核试验委员会，《肺结核链霉素治疗：英国医学研究委员会的调查》，《英国医学杂志》，第2卷，第4582期（1948年）：769-782。

6. D.T. Campbell和J.C. Stanley，《实验和准实验研究设计》，芝加哥：兰德·麦克纳利，1963年，第3页。

7. Esther Thorson、Robert H. Wicks和Glenn Leshner，《新闻学和大众传播研究中的实验方法论》，《新闻学与大众传播季刊》，第89卷，第1期（2012年）：112-124。

8.R. McDermott，《国际关系中实验研究的新方 向》，《国际研究季刊》，第55卷，第2（2011年）：503-520。

“Exploring the Promise of Experimentation in Political Science: Micro-Foundational Insights and

9. Ulrich Hamenstädt，《教授实验政治科学：

9. Matthew C. Makel等，《特殊教育研究的复制：必要但过于罕见》，《康复与特殊教育》，第37卷，第3期（2016年）：1-8。

10. Rebecca B. Morton和Kenneth C. Williams，《实验政治科学与因果关系研究：从自然到实验室》，纽约：剑桥大学出版社，2010年：5-6，16-17；Gerry Stoker，《探索政治科学中实验的承诺：微基础洞见和政策的相关性》，《政治研究》，第58期（2010年）：300-319；Cengiz Erisen、Elif Erisen和Binnur Ozkececi-Taner，《政治心理学研究方法》，《土耳其研究》，第13卷，第1期（2013年）：13-33；James N. Druckman等，《政治科学中实验研究的增长和发展》，《美国政治科学评论》，第100卷，第4期（2006年）：627-635。

11. Ulrich Hamenstädt，《教授实验政治科学：来自方法研讨会的经验》，《欧洲政治科学》，第11卷，第1期（2012年）：114-127；Erisen、Erisen和Ozkececi-Taner，《政治心理学研究方法》；James N. Druckman和Arthur Lipia，《进行政治实验》，《科学》，第335卷，第6期（2012年3月9日）：1177-1179；Luke Keele、Corrine McConnaughy和Ismail White，《增强实验者工具箱：内部效度的统计估计》，《美国政治科学杂志》，第56卷，第2期（2012年）：484-499；Sharon Crasnow，《政治科学中的自然实验和多元主义》，《社会科学哲学》，第45卷，第4/5期（2015年）：424-441。

12. Yair Levy、Timothy J. Ellis和Eli Cohen，《初学者研究人员的实验和非实验指南》。Yair Levy、Timothy J. Ellis和Eli Cohen，《初学者研究人员的实验和非实验指南》。

Chapter 1 ■ Discovering Cause and Effect **21**

151–161《信息系统研究中的准实验研究》，《跨学科信息、知识和管理学期刊》第6卷（2011年）：151-161。

13. Susan D. Hyde，《国际关系实验：实验室、调查和实地研究》，《政治科学年度评论》第18卷（2015年）：403-424。Hamenstädt，《教授实验政治学》。

14. Erisen、Erisen和Ozkececi-Taner，《政治心理学研究方法》。

15. William J. Reid、Bonnie Davis Kenaley和Julanne Colvin，《某些干预比其他干预更有效吗？一项比较社会工作实验综述》，《社会工作研究》第28卷，第2期（2004年）：71-81。

16. Joshua B. Plavnick和Summer J. Ferreri，《教育研究中的单一案例实验设计：教学与学习因果分析方法论》，《教育心理学评论》第25卷（2013年）：549-569。

17. William R. Shadish、Thomas D. Cook和Donald T. Campbell，《推广因果推断的实验和准实验设计》（加利福尼亚州贝尔蒙特：Wadsworth Cengage Learning，2002年）。

18. K. Imai、G. King和E. Stuart，《关于实验研究者和观察研究者对因果推断的误解》，《皇家统计学会A系列杂志》第171卷（2008年）：481-502。

19. Gustav Fischer，《鸟类学月报》，年份44，第2期（1936年）；Gustav Fischer，《德国市镇统计年鉴》，年份48，第1期（1940年）：27-33，http://pignottia.faculty.mjc.edu/math134/classnotes/storks.pdf。

20. http://faculty.vassar.edu/lowry/ch3pt2.html

21. Carter Hay等，《黑匣子内部：识别介入实验对青少年影响的中介变量》，《犯罪与犯罪学》第61卷，第2期（2015年）：243-270。

22. Renita Coleman，《视觉效果对道德推理的影响：记者在道德决策中照片的价值是多少？》，《新闻与传媒季刊》第83卷，第4期（2006年）：835-850。

23. H. Denis Wu和Renita Coleman，《情感对政治判断的影响：比较候选人属性和议题一致性的影响》，《新闻与传媒季刊》第91卷，第3期（2014年）：530-543。.

1. Daniel Stevens，《情调与信息：解释负面政治广告的影响》，《政治营销杂志》第11卷，第4期（2012年）：322-352。
2. Arthur Asa Berger，《符号学与电视》，收录于《理解电视：关于电视作为社会和文化力量的论文集》，编者R. R. Adler（纽约：普雷格出版社，1981年）：91-114。
3. 赵娜等，《面孔吸引力在建立信任中的作用：来自隐性和显性反应测量的证据》，《社会行为与人格：国际期刊》第43卷，第5期（2015年）：855-866；Leslie A. Zebrowitz、Robert G. Franklin Jr.和Rocco Palumbo，《生病的选民支持有吸引力的国会候选人》，《进化心理学》第13卷，第1期（2015年）：16-28。
4. Amy E. Lerman、Katherine T. McCabe和Meredith L. Sadin，《政治意识形态、肤色和候选人评价的心理学》，《公共舆论季刊》第79卷，第1期（2015年）：53-90；Sarah Reckhow、Matt Grossmann和Benjamin C. Evans，《政策线索和态度中的意识形态：对特许学校的态度的证据》，《政策研究期刊》第43卷，第2期（2015年）：207-227；Jonathon Schuldt和Adam Pearson，《种族和族裔在气候变化极化中的作用：来自美国全国调查实验的证据》，《气候变化》第136卷，第3/4期（2016年）：495-505。
5. Stoker，《探索政治科学实验的潜力》。
6. Kevin J. Mullinix等，《调查实验的普遍适用性》，《实验政治科学杂志》第2卷（2015年）：109-138。
7. J. R. Rest和Darcia Narvaez编辑，《职业中的道德发展：心理学与应用伦理学》（新泽西希尔斯代尔：厄尔鲍姆出版社，1994年）。
8. James. R. Rest，《道德》，收录于《儿童心理学手册》，第III卷：认知发展，编者P. H. Mussen（纽约：威利出版社，1983年）。
9. Stephen Thoma，《伦理问题理解和偏好中的性别差异估计》，《发展评论》第6卷，第2期（1986年）。
10. A. J. De Craen等，《医学中的安慰剂和安慰剂效应：历史回顾》，《皇家医学协会杂志》第92卷，第10期（1999年）：511-515。
11. 同上。

35. 同上；David B. Elliott，《安慰剂效应：使用它是否不道德，不使用它是否不道德？》，《眼科与生理光 学》第36卷，第5期（2016年）：513-518。

36. Elliott，《安慰剂效应》。

37. H. K. Beecher，《强大的安慰剂》，《美国医学会杂志》第159卷，第17期（1955年）：1602-1606。

38. De Craen等，《医学中的安慰剂和安慰剂效应》。

39. Beecher，《强大的安慰剂》。

40. De Craen等，《医学中的安慰剂和安慰剂效应》。

41. H. K. Beecher，《手术作为安慰剂：偏见的定量研究》，《美国医学会杂志》第176卷（1961年）：1102-1107。

42. Beecher，《强大的安慰剂》。

43. P. C. Goetzsche，《安慰剂中有逻辑吗？》，《柳叶刀》第344卷，第8927期（1994年）：904。

44. Alan Bryman，《安慰剂》，收录于《社会科学研究方法的Sage百科全书》，编者Michael S. Lewis-Beck、Alan Bryman和Tim Futing Liao（加利福尼亚州千橡：Sage，2011年）：825。

45. 同上。

46. Aaron M. Garvey、Frank Germann和Lisa E. Bolton，《性能品牌安慰剂：品牌如何提升表现，消费者却获得赞誉》，《消费者研究杂志》第42卷，第6期（2016年）：931-951。

47. Michael A. Kuhn、Peter Kuhn和Marie Claire Villeval，《决策环境对时间跨度金融选择的影响：资源耗尽模型有多相关？》，《经济行为与组织杂志》第137期（2017年）：72-89。

48. Yann Cornil、Pierre Chandon和Aradhna Krishna，《红牛给伏特加带来翅膀吗？市场营销标签对感知中毒和风险态度与行为的安慰剂效应》，《消费者心理学杂志》（爱思唯尔科学）第27卷，第4期（2017年）：456-465。

49. Carlos Alberto Alves、Evandro Luiz Lopes和José Mauro da Costa Hernandez，《它让我感觉很好：关于品牌产生的安慰剂效应的实验研究》，《国际消费者营销杂志》第29卷，第4期（2017年）：223-238。

50. 卡洛斯·阿尔维斯（Carlos Alberto Alves）、埃万德罗·路易斯·洛佩斯（Evandro Luiz Lopes）和何塞·毛罗·达·科斯塔·埃尔南德斯（José Mauro da Costa Hernandez），《让我感觉很好：品牌产生的安慰剂效应的实验研究》，《国际消费者营销杂志》29卷，第4期（2017年）：223-238。

50. Gwen Sharp，《安慰剂效应》，《社会页面》，2011年3月10日，https://thesocietypages.org/socimages/2011/03/10/the-placebo-effect/。

51. Mike Schmierbach和Anne Oeldorf-Hirsch，《有只小鸟告诉我，所以我不相信：Twitter，

**22** Designing Experiments for the Social Sciences

52. Jens Grober, Ernesto Reuben, and Agnieszka Tymula, “政治条件交换协议：一项实验研究,” 《美国政治科学杂志》 57, no. 3 (2013): 582–597.

53. Shirley S. Ho and Douglas M. McLeod, “面对面和计算机中介通信中的社会心理影响对意见表达的影响,” 《传播研究》 35, no. 2 (2008): 190–207.

54. Julio J. Elias, Nicola Lacetera, and Mario Macis, “市场与道德：一项实验调查研究,” 《公共科学图书馆》 10, no. 6 (2015): 1–13.

55. Pazit Ben-Nun Bloom, “政治中的厌恶、伤害和道德,” 《政治心理学》 35, no. 4 (2014): 495–513.

56. 例如，参见William Minozzi等人的“在线市政厅中国会成员的实质性、归因和行为说服的实地实验证据,” 《国家科学院院刊》 112, no. 13 (2015): 3937–3942.

57. R. Barker Bausell, “进行有意义的实验：成为科学家的40个步骤,” (千橡市，加利福尼亚州：Sage, 1994).

58. FoxNews.com, “参议院报告发现科学基金研究浪费了数十亿美元,” Foxnews.com, http://www.foxnews.com/politics/2011/05/26/senate-report-finds-billions-waste-science-foundation-studies/.

59. Nicole Neil and Emily A. Jones, “研究治疗强度：来自两项初步研究的教训,” 《行为教育学杂志》 24 (2015): 51–73.

60. Renita Coleman, “色盲：非洲裔美国人对新闻伦理困境的种族和伦理推理,” 《新闻学与大众传播季刊》 88, no. 2 (2011夏季): 337–351.

61. Sean Aday, “新闻的框架效应：倡导主义与客观主义框架的实验测试,” 《新闻学与大众传播季刊》 83, no. 4 (2006): 767–784.

62. Elias, Lacetera, and Macis, “市场与道德,” 2.

63. Grober, Reuben, and Tymula, “政治条件交换协议,” 582–583.

2

# 伦理和历史上著名实验

*If如果我比别人看得更远，那是因为我站在巨人的肩膀上。*

*- 艾萨克·牛顿*

|  |
| --- |
|  |
| 学习目标 |
| • 定位当今实验研究中所使用的方法、技术和概念的历史起源。  • 阐明创造力在解决研究问题中的重要性。  • 总结实验研究的无意后果，作为制定伦理准则和伦理审查委员会的基础。  • 解释在实验中使用欺骗的有害影响。  • 描述如何在实验的背景下道德地使用欺骗。 |

这一章简要偏离了实际的操作方法，为了提供一些背景来理解社会科学实验中当前问题、概念和技术的作用。这个简要历史将通过讲述一些最著名和臭名昭著的实验的故事来说明实验方法论的发展。

T

**23**

过去的历史可以追溯到一些方法、技术和概念的起源，例如随机化或将被试分配到不同组别的过程，以及为什么实验研究者使用诸如分割区块之类的术语。一些故事突出了回答问题的巧妙方式，鼓励研究人员在自己的研究中富有创造力地思考。创造力在研究中的重要性不仅在于追求新颖性，而且在于以最佳方式回答问题。其他的例子展示了保护参与者免受伤害的重要性，以及它们如何帮助推动伦理审查委员会（IRB）的发展，这些委员会负责确保研究人员遵守伦理准则。其中许多故事非常有趣，有些甚至被改编成了面向大众的电影。更重要的是，理解历史根源可以更好地掌握当代的实践。其他的例子展示了实验研究结果如何影响现实世界的问题，例如著名的研究探讨了为什么好人会做坏事。本章介绍了一些“巨人”的故事，正如艾萨克·牛顿的引言所说，实验方法论正是建立在这些巨人的肩膀上。

实验设计的历史中充满了引人入胜的研究，这些研究常常被炒作为最邪恶、令人毛骨悚然、离奇或者出乎意料的研究。那些无法再进行的研究也是“疯狂科学”类别中的热门主题。本章将采取不同的方法，不再详述这些实验产生的发现或理论，而是着重介绍随机分配、对照组以及同伴（Confederate）的使用等重要发现的发展和早期运用。重要的是，本章还将包括伦理准则的发展和IRB的形成过程（第11章将更详细地讨论IRB和其他伦理问题，这里不完全涉及）。许多这样的研究已经通过流行传说广为人知；在这些情况下，我尽量避免重复显而易见的内容，而是突出一些不太熟悉的方面。本章并不旨在详尽无遗地涵盖所有重要的实验，必然地会有很多被遗漏的实验。特别是，传统的历史倾向于忽视女性和有色人种的实验者。本章在“更多内容……”框2.1中涉及了一些被遗忘的人物，即女性的贡献。许多历史学家将实验的发展追溯到古希腊或其他时期。在介绍了一项非常早期而重要的医学实验之后，我将本章的剩余部分献给了更现代的社会科学实验。

#### 坏血病研究

早期的一些实验是在1747年针对坏血病进行的。尽管如今罕见，但在船上特别成问题，因为它会导致水手体弱贫血，皮肤出血，牙龈腐烂。船医詹姆斯·林德进行了一项最早的对照实验以寻找治疗方法。他选择了十二名患有坏血病的男子，尽量使他们“相似”。然后他将他们分成六组，每组两人。这六组人被分别给予不同的治疗方法，如柑橘水、醋、硫酸等。通过观察不同组的反应，林德发现柑橘水对于治疗坏血病效果最好。这项实验为坏血病的治疗方法奠定了基础，也标志着实验研究方法的发展。



### MORE ABOUT . . . BOX 2.1

Contributions of Women

尽管社会科学实验设计的早期先驱大多以男性为主，但也有许多重要的女性，包括有色人种女性。这些科学家在职业生涯中面临着各种障碍和歧视，包括缺乏奖学金、被拒绝进入研究生项目以及无法在允许研究和发表论文的学术职位上就业。一些女性完成了博士学位的研究工作，包括毕业论文，但被拒绝授予学位。男性获得了一些女性贡献的认可。一些女性与丈夫合作，而反亲戚主义政策阻止了她们的雇佣。种族歧视给少数族裔女性带来了更大的负担。

很少有历史包括女性，主要是指她们的理论贡献而不是她们在实验设计方法论上的进展，而这正是本书的主题。接下来是三位女性的简要介绍，其中包括一位非洲裔美国人，她们对实验方法论做出了贡献。更多的女性当然值得被认可。

玛丽·惠顿·卡尔金斯是实验心理学的先驱之一，她因发明了成对联想任务而

闻名，这是一种使用配对数字和颜色进行记忆测试的方法。尽管她完成了

博士学位的所有要求，但哈佛大学拒绝授予她学位，原因是她是女性。

耶鲁大学和密歇根大学向她提供了录取机会，但她拒绝了，因为这些学校

没有实验室，而哈佛有。对女性学者来说，在女子学院找到学术职位更

容易，所以卡尔金斯去了韦尔斯利学院，并在那里建立了一个实验室。

尽管没有博士学位，卡尔金斯在科学期刊上发表了四本书和100多篇论文。

她的成就使得哥伦比亚大学和史密斯学院授予她荣誉博士学位。尽管获

得了这种认可，以及哈佛签署的一份请愿书上有十三位重要的校友签名，

但她在1927年再次被拒绝授予学位。

卡尔金斯在本章中与另一位实验研究者约瑟夫·贾斯特罗交集，贾斯特罗与

查尔斯·S·皮尔斯一起帮助确定了随机分配的好处。贾斯特罗发表了一项

研究，研究男性和女性在被要求迅速写出单词时产生的类型，得出结论

认为女性的单词是重复的、个体的和具体的，而男性的单词是建设性的、

有用的和抽象的。卡尔金斯批评了他的结论，指出环境、训练和女性

社会化的混杂效应。她一生都是妇女权益的倡导者，多次拒绝接受哈佛

哈佛德学位，因为她在哈佛的工作成果。

玛米·菲普斯·克拉克是一位非洲裔美国心理学家，因她在种族研究中



(Continued)

哥伦比亚大学。从哥伦比亚获得博士学位的第一位非洲裔美国男性是她的丈夫和研究伙伴肯尼思·B·克拉克。他始终认为是她提出了玩偶测试的想法。

玛米·克拉克在实验方法方面的贡献包括向儿童展示两个完全相同的玩偶，一个是黑人，一个是白人。问孩子哪个玩偶是坏的，哪个是好的，他们喜欢和哪个玩偶玩，并且哪个玩偶最像他们自己。许多黑人孩子认为黑人玩偶是坏的，几乎一半的孩子认为白人玩偶最像他们。在种族隔离的学校里，这种情况比集成学校更为明显。克拉克夫妇的实验成为证明种族隔离对儿童有害的重要证据，并对第一个大规模生产的黑人婴儿玩偶产生了影响。克拉克从未在学术界找到职位，而是在一个儿童之家担任研究员和临床心理学家，直到她和丈夫在哈莱姆开办了一家为少数族裔儿童提供测试和咨询服务的中心。

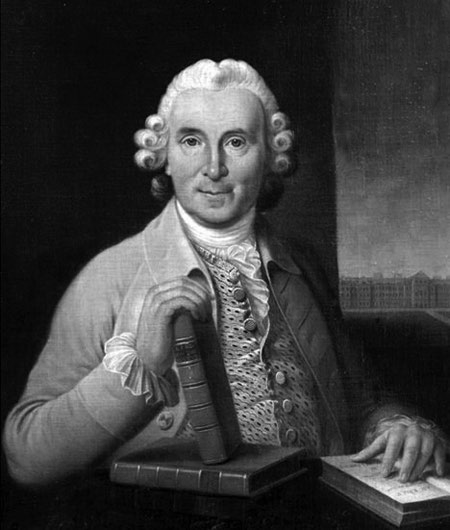
玛丽·埃因斯沃思因“陌生情境”评估技术而被人们铭记。使用这种方法的研究有助于心理学家了解儿童与照顾者的依恋关系。埃因斯沃思的方法是让研究人员透过单向镜观察孩子在大约三分钟的八个不同情景中的行为，这些情景是母亲和孩子来到研究人员的实验室，房间里有玩具。不同的情景包括陌生人进入房间并试图与孩子交朋友，母亲将孩子独自留给陌生人，母亲和陌生人离开孩子独自待着，陌生人再次进入安慰孩子，最后，陌生人抱起孩子。每15秒记录一次观察结果，使用1至7的评分标准。该评估具有很好的可靠性，即其他研究人员可以重现研究结果。作为实验室实验的典型特征，这种方法受到批评，认为它是人为的，缺乏生态效度，并且基于伦理原因对幼儿造成压力。

还有许多其他女性通过编写或合著其他测试和测量工具为实验方法论做出了贡献，包括格雷斯·肯特-罗桑诺夫的词语联想测试、弗洛伦斯·古登努夫的画人测试以及格雷斯·费尔纳尔德的品格测试，这些测试在道德发展等其他测试之前出现，等等。

然后，这些群组接受了不同的治疗，包括一夸脱苹果酒、硫酸、半品脱海水、一剂由大蒜、芥末和辣根混合而成的“一颗肉豆蔻大小”的药剂、醋，或者两个橙子和一个柠檬。林德在这项医学研究中设定了六个治疗条件，在当今社会科学中被认为是很多的，通常情况下会有两到四个条件。他只有很少的受试者数量，仅有十二人，每个治疗组只有两人。在医学研究中，小样本量比社会科学中更为常见，因为医学研究中通常存在着更显著的效应。这项研究中的另一个重要问题是如何将人们分配到不同的治疗条件中。林德认识到了这些群组中人们在重要因素上的差异，了解到人们固有的个体差异，例如年龄、体重和一般健康状况，可能会影响结果。为了得到不受这些外部因素混淆的治疗效果，他选择了在某些要求上尽可能相似的男性作为样本。在这里，他采用了后来被称为匹配策略的方法，将具有相似特征的男性进行配对。重要的是，他只操纵了这六个条件，同时保持其他一切（他能想到的）都保持不变。他的自变量——这六种不同的膳食补充剂，并非来自于理论，而是来自先前提出的治疗方法。然而，林德确实写到了关于坏血病的理论大多是从未见过该病症的研究者的猜测，并主张理论与实践相结合。他对其中一种治疗剂量的测量——肉豆蔻大小——并不十分精确；如今，这将以克或类似的单位进行测量。无论从任何标准来看，林德的研究都取得了巨大的成功。仅仅在六天之后，那些摄入橙子和柠檬的男性就恢复了健康。其他人也有所改善，但吃柑橘水果的人与其他人相比有了显著的恢复，这最终导致人们认识到维生素C才是起作用的物质。虽然没有进行显著性统计检验，但六天后的效果是显而易见的。

这项研究之所以重要，有几个原因：它认识到人们需要以一种不受个体差异影响的方式被分配到治疗组中，而林德通过精心选择具有相似特征的男性来实现了这一点。这发生在1747年。到了19世纪末，将受试者分配到条件中已经成为一个紧迫的问题，出现了不同的解决方案，尽管这些方案更多地应用于治疗方法而不是受试者，如下一个故事所示。林德的研究在任何标准下都是巨大的成功。仅仅经过六天，那些食用橙子和柠檬的男性就康复了。其他男性也有所改善，但食用柑橘水果的人相比其他人有了明显的恢复，最终导致人们认识到维生素C是起作用的因素。并没有进行统计学上的显著性检验，但六天内的效果是显而易见的。

这项研究之所以重要有几个原因：首先，它认识到人们需要以一种不受个体差异影响的方式分配到治疗组中，林德通过精心选择具有类似特征的男性成功地实现了这一点。这是在1747年。到了19世纪末，将受试者分配到条件中已成为一个热门话题，并出现了一种不同的解决方案，尽管这种解决方案是应用于治疗方法而不是受试者，这将在接下来的故事中说明。



詹姆斯·林德

#### 查尔斯·皮尔斯的贡献

查尔斯·桑德斯·皮尔斯和他的学生约瑟夫·贾斯特罗（Joseph Jastrow）据说是第一批使用随机分配的人。他们在1885年进行的一项研究中首次采用了随机分配，以观察人们是否能够通过感觉和观察来判断物体的重量。皮尔斯关注的理论构建是判断错误的源头，从而产生了诸如“刚刚能察觉的差异”之类的概念，这些概念在商业、营销和广告学者中得到了认可。在第一项研究中，皮尔斯始终从最重的物体开始和结束实验。在第二项研究中，他尝试将最重的物体与最轻的物体交替使用。最后，他描述了如何使用一副扑克牌进行随机分配为了随机分配重物的顺序，皮尔斯的结果与前两次试验完全不同，这显示了这种技术的重要性。他使用了一副扑克牌，简单地洗牌后抽出一张牌来决定下一个重物的顺序。如今，我们倾向于使用随机数生成器，但一副扑克牌仍然可以同样有效地使用。他提到这种方法有时会产生“长时间以一种特定的变化方式运行，这种变化有时会出现凭借机会”，但他指出这比被试知道没有这样的模式更可取。对于大多数人来说，“随机”意味着不可预测；但对于皮尔斯来说，随机意味着“在长期运行中，任何一个个体被选中的次数与其他个体一样频繁”。当以这种方式从较大的类别中抽取较小的样本时，皮尔斯说较小的群体会显示出与较大群体相同的特征。他称之为“归纳规则”。第7章将深入探讨社会科学中的随机分配；如今，我们更常常考虑将人们随机分配到不同的处理组，但正如皮尔斯和其他人的早期研究所示，随机化尽可能多地包括实验中的刺激是很重要的。

顺便提一下，为了展示皮尔斯的能力范围，他还因发展了符号学和象征学理论而闻名——这是一种经典理论，今天仍然在批评/文化研究中使用。

#### 罗纳德·费雪的茶

几十年后，罗纳德·艾尔默·费雪重新提出并推广了治疗的随机分配的想法，从而导致其被广泛使用。费雪是20世纪20年代和30年代伦敦外的一位农业科学家。他被认为是实验设计的发明者，包括今天仍在使用的许多统计学、概念和程序。例如，费雪开发了方差分析，其统计量F检验以他的名字命名，并且还提出了拉丁方阵的概念，一种用于对治疗进行排序以补偿系统误差或控制意外变异的程序（关于这一点将在第7章中详细介绍）。费雪还被认为提出了0.05和0.01的概率值作为统计显著性的标准，这是科学家判断某个事物是否偏离机会的方法。0.05的p值意味着“结果错误的概率为二十分之一”。费雪的工作是在英国罗瑟姆斯特德农业实验站上对各种肥料进行测试。因此，许多今天在实验中仍然使用的术语都具有农业基础，例如分割剧情设计，实际上是指将田地分割成不同部分，以便每个部分可以接受不同的处理。费雪到达实验站时，一直以来的惯例是每年测试一种不同的肥料。但是费雪意识到，每年带来了降雨量、温度、杂草生长、排水和其他因素的固有差异。当他意识到无法将肥料的效果与这些无法控制的其他条件分离时，他创造了混杂这个术语。为了确定结果是由肥料还是其他因素引起的，他决定将所有的处理方式都纳入到同一个实验中，对他来说，这些处理方式是许多不同种类的肥料。通过将田地划分为小区，将小区划分为不同的行，并给每一行分配不同的处理方式，他推理出每年的差异将适用于所有的处理方式，有效地“控制”了这些混杂因素。此外，费雪还被认为推广了随机分配的方法。这也源于他在努力控制各种条件时意识到，将不同的肥料有系统地分配给田地不能排除与土壤和田地本身相关的混杂因素。为了解决这个问题，他随机分配田地接受不同的处理方式。在当今社会科学中，我们认为随机分配适用于被分配到不同处理方式的受试者；对于费雪来说，田地就是他的受试者。从随机分配的意义上讲，“随机”意味着按照机会程序，例如抛硬币，确保每个参与者有平等的机会被分配到任何一个治疗或对照组。这有助于确保任何系统性差异在不同组之间平均分布，从而可以将任何差异归因于治疗，而不是组内人员固有的差异。第7章将详细讨论随机分配及其实现的更多细节。费雪进行的一个较为引人注目的随机分配实验是现在著名的“品茶女士实验”。它还说明了零假设的概念，即认为接受治疗和不接受治疗的人之间不会有差异。故事是这样的，一个声称她可以通过品尝告诉奶先加还是茶先加的女士被给予了四杯先加茶的杯子和四杯先加奶的杯子，其顺序是随机的。零假设是她无法将它们区分开来。有人全部正确猜测的概率是七十分之一；据说这个研究中的女士全部都答对了。这个品茶实验据说是一个夏日的娱乐活动，并不是一项具有出版结果的科学调查。费雪在他的著作《实验设计》的第2章中描述了这个实验。他没有说明是否实际进行了实验或给出了结果，所以其中的许多内容都是民间传说。在费雪之前，已经进行了几百年的实验，但它们是非常个人化的，每个实验者的方法都有所不同。费雪的书是第一本系统总结如何进行实验的著作。

A person with a beard and mustache

Description automatically generated

罗纳德·艾尔默·费雪

Wikimedia Commons

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

i作为一名遗传学家，费雪对进化和优生学（选择性育种可以改善人类种群的理念）也很感兴趣。优生学的讨论可以追溯到古希腊和古罗马时期，在费雪所处的时代，它是一个备受尊重的科学话题，得到了包括乔治·伯纳德·肖和亚历山大·格雷厄姆·贝尔在内的许多人的支持。费雪还帮助创立了剑桥大学的一个优生学社团。然而，由于与纳粹德国的种族灭绝政策相关联，这个理念在之后彻底被抹黑。（格雷戈里·科赫兰和亨利·哈普丁，《一万年的爆炸：文明如何加速人类进化》[纽约：基础图书出版社，2009]；《大英百科全书》，罗纳德·艾尔默·费雪，https://www.britannica.com/biography/Ronald-Aylmer-Fisher，于2018年3月30日访问；《著名科学家》，https://www.famousscientists.org/ronald-fisher/，于2018年3月15日访问。）

#### B. F. 斯金纳：小样本，高技术

费舍尔的原理在B.F.（缩写自Burrhus Frederic，但朋友们都叫他Fred47）斯金纳开始研究老鼠的行为时才开始受到欢迎。48斯金纳以创建斯金纳盒子而闻名，这是他用来训练老鼠推动杠杆以获取食物或停止电击的装置，他在开发他的操作性条件反射理论时使用了这个设备。在这里不对他的理论或研究进行详述，但基本的前提是动物学会对刺激做出反应，而强化某种行为会使动物重复这种行为。尽管斯金纳的想法有时会引起争议并被误解，但他的工作和理论仍然在激励着今天的研究，包括在人类心理学、动物行为、教育、营销、健康、社会工作、运动训练和其他领域。49这里的讨论是专门针对他对实验研究方法的观点，因为这就是本书的目的。

斯金纳被认为是开发了一种他称之为行为的实验分析的心理学实验研究。50这与当时大多数心理学实验所使用的演绎方法形成了鲜明的对比，该方法首先形成一个假设，然后测试看看它能否被证伪——这种方法仍然是现在最常用的方法。斯金纳的方法是归纳的和以数据为驱动的，特征是观察和实证测量行为。他从未使用过正式的假设，他说：“如果我参与了实验设计，那只是为了完成或扩展已经观察到的某种秩序的证据。”51他在他的书《生物的行为》52中详细地发布了他的方法，这比当时的常规更详细。53

他的大部分研究都是以个案进行的，一次只进行一个主题或在小组中进行；在《生物的行为》中，大部分研究都使用了四只老鼠。54他为他使用小样本的做法进行了辩护，并描述了它们可能比大样本更好的情况55——例如，当研究招募大量同质性受试者困难的群体时，如特殊教育学生、精神病患者或残疾人。58今天，这种方法被称为单一案例实验设计，仍在用于此类人口，尤其是在临床工作中。59

斯金纳研究大群体的一个困难原因是每个受试者都需要一个斯金纳盒子的限制。他的实验室里只有四个这样的盒子（他更喜欢称它们为“操作性室”）。60他描述了他曾经如何获得一项资助，用以制造足够的盒子同时测量24只老鼠，并将数据集中在一个平均性能曲线中。他和他的同事在一项研究中能够运行95只老鼠，斯金纳写道“使用大群体的动物的可能性大大提高了以前报告的方法，因为提供了显著性的测试，并且在单一案例中不明显的行为特性可能更容易被检测到。”61但是，他最常见的做法是运行个别受试者或样本太小以至于无法进行任何统计分析。他依赖功能分析来确定干预是否有效。例如，当八只鸽子中有六只执行了他们被训练进行的行为时，“所产生的反应定义得如此清楚，以至于两个观察者可以完全一致地计算实例。”62他觉得当真正需要的是更严格的控制和技术时，统计被过分强调。63他将“控制你的条件，你将看到秩序”这个洞见归功于巴甫洛夫。64

斯金纳还将实验方法从研究者的观察推进到更客观的测量工具。斯金纳从小就是一个发明家，当他还是个男孩的时候，他就建了一个设备提醒自己把睡衣挂起来。65为了配合斯金纳盒子的工作，他设计了一个“累积记录器”，用笔和纸卷自动记录当老鼠按下杠杆或鸽子啄键时的情况。其他研究者广泛采用了反应率的使用。66他继续设计了他所称的“小工具”或“设备”来进行许多其他用途。67例如，他使用秒表记录老鼠跑动和发出声音的行为。68另一个设备使用了一个绕在纺锤上的细线来记录数据的曲线。“我知道科学大量利用曲线，尽管根据我所了解的，在聚集图（polygram）上的小纹理（pips）使用得非常少。”69他在谈到他以前的测量方法时写道。“结果显示，曲线揭示了反应速率以及该速率的变化中的事物，这些事物在其他情况下肯定会被忽略。”70响应曲线在许多学科中仍然是一个重要的原则。71他另一项发明是改装一台药剂师制药机，制作出大小统一的老鼠食粮颗粒，解决了现有老鼠食品大小不受控制的问题。当他意识到为了强化每一个行为，他需要制作多少老鼠食粮时，他开始每分钟只强化一次，这启动了一项关于“周期性强化”的新的研究项目。72

这些实际问题驱动了他的大部分研究，他开始将这些问题视为意外的幸运，而不是烦恼。例如，当老鼠食粮输送装置卡住时，他写道：“我最初将此视为一个缺陷并急于修复”它。73然后，他意识到看似挫折的事情实际上提供了发展“消退曲线”的机会。“毫不夸张地说，一些最有趣和令人惊讶的结果首次出现是因为类似的意外。”74

斯金纳将科学称为一个“混乱且偶然的过程”，没有“明确定义的开始和结束”。75《生物的行为》76描绘了科学的混乱性和偶然发现的角色，也就是在寻找一件事物时发现另一件事物。（关于现代的例子，参见2.2研究聚光灯的“混乱”研究。）在斯金纳的例子中，他乐于看到这种情况。他用自己的话说：“这里有一个科学方法学家们并未正式认可的第一原则：当你遇到一些有趣的东西，把其他所有事情都放下，去研究它。”77

这仍然是今天的好建议。

ii他为鸽子和婴儿开发了另一个盒子；一名学生为狗开发了一个，后来受斯金纳启发的另一名学生开发了一个人类的斯金纳盒——今天的学习小隔间。参见 Rutherford 2007，2003.56，57。

t



B. F. 斯金纳

### 研究聚光灯 2.2

“混乱”的研究示例

巴巴德、皮尔和贝纳约恩（Babad, E., E. Peer, and Y. Benayoun）在2012年的《应用社会心理学杂志》（Journal of Applied Social Psychology）发表的研究论文《一个情境中是否可能存在多重偏见？从媒体偏见研究中寻找证据》（Can Multiple Biases Occur in a Single Situation? Evidence From Media Bias Research）是研究混乱性的当代案例。这篇文章中罕见地坦率解释了研究如何偏离轨道、产生了意想不到的结果和违反直觉的发现。然而，这种情况比它看起来的频率更高，发现在论文中能找到这样坦诚的讨论令人耳目一新，也让人感到安心。这篇文章还展示了计划之外可能出现的意外结果。

这篇论文的第一句话承认了这个想法是由于实验研究的意外结果而出现的。这项研究涉及影响人们判断和行为的认知和情绪偏见。在这些研究人员的领域中，这意味着当电视记者采访政治候选人时，他们的非语言偏见（NV行为）是无意识的。这些研究人员研究的问题是实验设计中一些最好的技术导致的，具体来说，设计出不受干扰影响的实验与现实生活中发生的情况不同。此外，实验研究者倾向于一次研究一种偏见，而不像在日常生活的复杂性中，同时存在多种偏见的情况更有可能。作者们承认，最初他们没有想到这个想法，也没有发表的研究指出可能存在多重偏见的可能性。他们说：“我们必须承认最初我们没有构思出多重偏见的概念，并且也没有研究表明可能存在多重偏见的可能性。”78

在得出两种不同的偏见同时起作用的结论之前，研究人员进行了自己研究的艰难复制，涉及了七个不同的样本。这篇论文报告了这些研究的元分析结果。在研究结束时，他们坦率地指出哪些发现是出乎意料的、违反直觉的、令人惊讶和令人失望的。他们承认当发现难以解释时。例如：“在这项研究中，光环效应并不是真正预期的，因为实验并没有被设计成产生光环效应...最令人意外和违反直觉的结果，也许具有最广泛的概念和应用意义，是两种偏见现象的独立性，这一点通过拒绝替代光环效应假设得到了证明...很难解释这种效应，确实是令人费解的。”79

这些研究人员并没有试图掩盖和伪装，他们展示了实验研究的困难和混乱，这种情况比发表的文章给研究人员带来的印象更常见。

#### 斯坦利·米尔格拉姆震惊世界

斯坦利·米尔格拉姆的研究是一些最为人熟知的，无论是学者还是普通大众都知道。俗称为“电击实验”，斯坦利·米尔格拉姆关于顺从的实验旨在探究人们为什么会服从权威人士的命令，即使这些命令违背了他们自己的良知。在20世纪60年代，

纳粹战犯审判正在进行中，因此对这个问题的兴趣非常浓厚。米尔格拉姆的实验因引发心理伤害和欺骗性而受到心理学家的批评。这些批评引发了对社会科学研究伦理的更广泛讨论。这些研究也被认为有助于建立IRB（伦理审查委员会），以确保研究不会伤害志愿参与的人（有关IRB的更多信息请参见第11章）。

虽然这些研究主要被记作关于伦理研究的警示故事，但其中还有其他方面适合作为实验设计教训的内容。米尔格拉姆的所有研究的主要结构是让一个“天真”的受试者（即不知道研究目的的人）“教授”另一个人一组词语配对。受试者被告知目标是观察惩罚（电击）是否能提高学习效果。在这种情况下，“学习者”是实验者的同谋或共谋者，他知道研究的目的，只是假装受到了电击。在那个时候，这样使用共谋者和欺骗的实验很受欢迎。米尔格拉姆在每个实验结束时进行了解释说明会，告诉受试者实验的真实目的，即他们实际上并没有伤害任何人，他们的反应是正常的。在发表了两篇文章并开始出现批评之后，米尔格拉姆开始报告他采取的措施来最小化对受试者的伤害以及他进行的事后调查以了解是否存在伤害。戴安娜·鲍姆林德（Diana Baumrind）驳回了米尔格拉姆的事后实验自我报告，这些报告显示超过80％的受试者对他们的欺骗性实验表示高兴，她说这些报告是“事后加上的”，“毕竟，如果自我报告可以被视为衡量实验条件影响的准确指标，我们完全可以摒弃实验操纵和行为测量。”如今，共谋者已经不再流行；然而，在涉及欺骗的实验中，解释说明仍然是最佳实践，尽管像米尔格拉姆的研究这样的实验今天很可能不会获得IRB的批准。然而，值得注意的是，在电视节目中，如ABC的秘密摄像系列节目《你会怎么做？》中，这种类型的“实验”仍然很受欢迎。当记者或娱乐制片人进行此类实验时，无需获得IRB批准。

第1章解释了因果关系的概念。米尔格拉姆感兴趣的效果是顺从，以最大电击的依赖变量来衡量，其范围从0到30.124。他所研究的原因是各种独立变量，包括受试者与受害者的距离、受试者是否能听到抗议和哭喊声、是否必须接触受害者、给予指示的人的权威性等等。

### 更多内容，第2.3个方框

欺骗

虽然社会科学实验中的欺骗可能不再像斯坦利·米尔格拉姆的电击实验那样引人注目，但它并不是过去的事情。它在不同学科中使用的方式各有不同；在主要协会的规范下，社会学和社会心理学中很常见，但在经济学中被禁止，尽管隐瞒信息不被视为欺骗，所以是被允许的。

欺骗被定义为有意提供虚假信息和隐瞒信息以误导他人。这并不妨碍研究人员隐瞒假设、实验条件或其他可能导致受试者改变行为的研究方面的信息。然而，如果隐瞒此类信息会导致受试者不同意参与，则超过了底线。

在可能导致身体疼痛或严重情绪困扰的研究中，禁止使用欺骗。指导方针要求研究人员在没有其他可行的方法或者研究价值得到充分证明时，不得使用欺骗。然而，正如鲍姆林德所说：“很容易让研究人员或同行评审委员会相信，一种巧妙的欺骗操纵的长期效益超过了被欺骗者的短期成本。”

欺骗性研究并不局限于斯坦利·米尔格拉姆的电击实验。在该实验之后，菲利普·辛巴达罗（Philip Zimbardo）在考虑使用其他方法和雇佣客观观察员来监督研究时进行了反思。然而，在随后的几年里，辛巴达罗和同事们使用催眠诱发部分失聪，导致受试者出现偏执症状，对受试者提供错误信息并隐瞒研究目的和他们将要经历的事情。就像监狱实验一样，这项研究和另一项涉及误导受试者的研究都获得了IRB的批准。1977年，斯坦利·米尔格拉姆的电击研究被复制，并对年龄只有六岁的儿童进行了实验。这项研究是在一个没有IRB的国家进行的，但获得了系主任和其他教师、校长的批准。研究报告称参与的儿童表现出“大声紧张的笑声，咬嘴唇，颤抖”。在文章的结尾，作者们说，“确实令人惊讶的是，相对较少的社会心理学家对米尔格拉姆在顺从方面的开创性工作进行了后续研究。”这项研究当时和现在都发表在顶级社会心理学期刊上，该期刊也发表了米尔格拉姆的另一篇论文。这并不意味着这些研究人员是不道德的，甚至是无视问题的，而是指出检测欺骗问题可能比人们想象的更困难。

内容分析显示了欺骗性研究的发展轨迹。从1921年开始，在领先的社会心理学期刊《人格与社会心理学杂志》中很少有欺骗，直到20世纪30年代。随着实验方法的兴起，它逐渐增加，尤其是在20世纪50年代至70年代之间。从20世纪80年代到1994年，欺骗性研究有所减少。研究发现，电脑和其他虚假设备部分取代了同谋者以误导受试者。在1969年之后，误导同意和对受试者的错误反馈有所减少，但在1992年又有所增加，研究主题的变化解释了在1978年至1986年间欺骗性研究的下降。对所检查的文章，发现其中高达66％存在欺骗性实践。

（接续）

研究显示类似的结果。在1971年至1974年间，54%的心理学实验使用了欺骗，比1948年的18%和1963年的37%有所增加。最近的数据显示，在社会心理学研究中，约有33%的研究存在欺骗性。欺骗可能对受试者造成伤害是显而易见的。除了身体疼痛或心理困扰之外，欺骗还会损害研究对象的自主权，违反他们选择参与的权利。

但是，欺骗性研究还具有对研究行业整体产生影响的潜在危害。正如许多社会科学家指出的，参与研究的自愿性已经在下降。这在很大程度上归因于公众对研究的怀疑，这是基于研究在历史上造成伤害、欺骗和滥用权力的情况。这不仅仅影响了召集受试者的能力，还可能削弱公众对研究结果的信任和自信，以及整个学术界的声誉。社会变得不愿意支持研究。此外，一些实验者使用欺骗的意愿也有可能引导其他人也这样做，从而削弱研究人员自身的道德机构性。对整个社会的伤害包括整体信任的损失和越来越多的怀疑，因为违背了讲真话和信守承诺的社会契约。

以合乎伦理使用欺骗的最常见原因是需要控制需求特征，即受试者改变自己的行为以适应研究的目标的倾向。鲍姆林德认为，使用欺骗会促使其本来打算避免的事情发生：因为人们知道欺骗在实验中经常被使用，受试者不再天真，也就是他们不再相信研究是真实的。与此同时，其他研究发现，欺骗不会影响研究的外部有效性。

除了完全避免使用欺骗之外，合乎伦理使用欺骗的最常见技巧包括：

• 向受试者提供知情同意书，其中包括研究的真实目的，而不揭示假设，并描述研究的程序和他们将要做的事情。

• 如果有任何信息被隐瞒，在研究结束后立即向受试者进行解释说明。同时，应该询问受试者是否同意推迟接收所有信息直到解释说明会。然而，要认识到解释说明无法消除伤害。

• 允许受试者在任何时候退出研究，包括在解释说明后撤回他们的数据。

实验者应该意识到，他们可能倾向于选择欺骗性方法，因为这比非欺骗性方法更容易。相反，研究人员应该考虑使用非欺骗性的方法，例如自然实验，其中不涉及操纵，或者使用不显眼的工具，使受试者无法改变他们的反应，例如心率的心理测量或反应延迟，用于测量一个人对刺激的反应时间。自然实验在第4章中有详细介绍，不显眼的工具在第10章中有介绍。关于伦理和

在第11章的IRB（伦理审查委员会）背景下讨论欺骗。对于非欺骗性实验，还提出了其他建议，包括让研究人员成为自己的研究对象，研究他们的动机和行为，或者使用“替代性受试者”，即告诉人们关于研究目的和程序，看看他们是否接受。

最后，我的建议是遵循黄金法则，即待人如己。将自己置身于研究对象的角度，问问自己是否会对被欺骗、被谎言欺骗或发现自己的痛苦经历有任何异议。考虑后果，包括如果你的雇主、家人或朋友发现你宁愿保密的事情，那会是什么样子。



正如米尔格拉姆所写：“该研究的关键在于系统地变化被认为会改变对实验命令服从程度的因素。”他总共进行了十八个不同的实验，分别发表在国际期刊文章中，并在他的著作《对权威的顺从》中报道了整个系列实验。

第1章还介绍了“为什么”问题，即解释为什么一个实验很重要。米尔格拉姆通过引用他的研究可以揭示的实际问题回答了“为什么”问题，从战争罪行到员工服从上司。他还想出了一个聪明的办法来回应对仅确认我们已知信息的研究提出的批评，而不是揭示新的或意外发现的研究。他设计了一个程序来确定他的研究结果是否显而易见和符合直觉，即进行了一项单独的研究，邀请受试者参加关于他研究主题的讲座，并描述了实验而不透露结果。讲座结束后，受试者被问及他们自己会怎么做以及他们认为其他人会怎么做的几个问题。所有110名受试者都表示他们会在某个时候拒绝给予电击，并且几乎所有其他受试者也会如此。而在实验中的情况恰恰相反，65%的受试者一直进行到最高水平的电击。

因此，米尔格拉姆能够显示他的结果并不是显而易见的，实际上与人们预期的相反。

就像林德在他的坏血病实验中所做的那样，米尔格拉姆运用了理论和观察相结合的方法来测试服从的原因。例如，他在第8个实验中使用女性受试者，就引用了关于女性的矛盾理论作为原因。尽管在今天看来有些过时，但当时的一些理论和证据指出，女性比男性更服从和更少攻击性，这应该导致她们比男性更频繁地给予电击。而矛盾的理论和证据则表明，女性比男性更具共情能力，这预示着她们会比男性更少给予电击。实际上，女性和男性在电击实验中的表现大致相同，但在事后的访谈中，她们对给予电击的原因使用了不同的表述方式。第3章将涵盖理论对实验的重要性以及观察和经验如何推动实验问题的产生。

米尔格拉姆还在他的研究中讨论了抽样和参与者的问题。他明确拒绝使用本科生作为受试者，尽管这在今天相当普遍，但也引发了激烈的辩论。他指出，使用学生会更容易，因为他们可能已经从已经参与过研究的其他人那里听说过这些研究，并且他们太同质化，年龄、智力和对心理学研究的了解都相似。相反，他希望从不同社会阶层背景的广泛个体中选择。因此，他在报纸上刊登广告，从周围社区招募参与者，并以4.5美元的现金激励来吸引他们。米尔格拉姆在书中写到，每个条件下他都使用了四十名受试者进行所有的研究，但没有提到他是如何确定这个数字的。今天，会使用统计学中的功效分析（power analysis）来确定一个特定研究需要多少受试者。这将在本书的第8章中介绍。他也没有随机分配受试者到不同条件，这是为了确保等价性的特征，已经被其他人发现，但主要用于排序处理方式而不是分配受试者。相反，米尔格拉姆使用了一种类似于林德在坏血病研究中使用的配对技术。他提到平衡每个条件中的受试者在年龄和职业上的分布很重要，所以他将他们分成熟练工人和非熟练工人的类别。

Iii 关于使用学生或不使用学生的问题，在所有实验结束后，米尔格拉姆将成年公民的结果与他在试点研究中使用的学生的结果进行了比较，并发现没有差异。《对权威的顺从：一个实验观点》（Obedience to Authority: An Experimental View），第170页，米尔格拉姆著

在所有的研究中，米尔格拉姆还通过使用受害者的抗议录音来控制混淆因素，确保每个受试者和每个电击水平的抗议声都是相同的。

米尔格拉姆的研究采用了多种测量方法，这是第10章的主题。他的主要依赖变量是行为性的，即受试者所施加的“电击”水平，范围从零到最高的30级，代表着惊人的450伏特。实际上，最低的电击水平是20伏特。他还测量了电击的持续时间，即受试者按下开关的时间和按住开关的时间。

其他的测量指标包括受试者在实验结束后通过问卷报告的自我报告，包括他们在研究过程中经历的冲突、紧张和紧张感，以及他们对受害者所感受到的疼痛程度的估计。米尔格拉姆还使用了多种回答选择，包括开放式问题、投射性测试和态度量表。所有这些内容将在第10章中详细介绍。在米尔格拉姆的前四个实验中，还询问了受试者对实验者、受试者和受害者对于一个人被迫接受电击的责任分配情况。问卷还询问了通常的人口统计变量，如政治党派、宗教信仰、教育水平以及在这种情况下的兵役年限，因为这在1960年代是常见的，并且被认为对于一个人是否能很好地遵守命令很重要。

此外，米尔格拉姆还记录了实验过程中受试者和实验者之间的对话，并且研究人员记录了自己的观察。除了个别访谈和实验后的小组讨论，这些定性的测量方法还与实验后的个别访谈和小组讨论一起使用，这些内容在第1章中讨论了三重性的重要特征。

在米尔格拉姆的研究中，我们还发现了试点研究的例子，这是第11章的一个主题，以及操纵检查，这是第9章的主题。试点研究在正式研究启动之前用于测试实验，以解决问题。在使用本科生进行试点研究时，米尔格拉姆发现受试者需要练习读给学习者的单词，因此在正式研究开始之前，他加入了一个训练环节，受试者读十对单词。此外，试点研究中的另一个重要经验教训是“声音反馈”。最初，米尔格拉姆让受试者在另一个房间给受害者施以电击，而受害者保持沉默。他发现受试者在给予电击时没有变化，每个人都进行到最高电击水平。由于结果没有变化，没有人违抗命令，因此无法确定违抗的因果机制。因此，米尔格拉姆引入了受害者的呼喊、抗议和拍打墙壁的声音，这导致受试者在不同的时间点停止电击，从而获得了第1章中讨论的变化。

第9章的一个主题是如何确定操纵是否被视为真实。例如，如果实验要求受试者阅读用生动、描述性语言或非生动、抽象风格编写的信息，那么就有必要知道得到生动写作的受试者是否认为它比得到非生动写作的受试者更生动。

为了确定这一点，需要进行操纵检查。在米尔格拉姆的研究中，他想知道受试者是否相信受害者实际上接受到了痛苦的电击，并且是否相信他们自己是施加电击的人。只有两个受试者在实验后的问卷中回答这些问题时选择了“否”，因此认为操纵有效。

For Review-No Commercial Use(2023)

Philip Zimbardo

Getty/Andrew H. Walker/Staff

#### 菲利普·辛巴多：在斯坦福地下室中唤起良知

斯坦利·米尔格拉姆的电击研究不久之后，菲利普·辛巴多进行了一项同样引人注目的研究。就个人经历而言，米尔格拉姆和辛巴多在1949年在布朗克斯的高中相遇过。他们在1960年作为助理教授重新联络，米尔格拉姆在耶鲁大学，辛巴多在纽约大学。巧合的是，辛巴多最初构建了一个实验。

在耶鲁大学的实验室用于他的电击研究之后，米尔格拉姆搬出了“优雅互动实验室”，而辛巴多在全国另一个地方的地下室进行了一项备受关注的研究，其中一个引人关注的问题是伦理问题。

1971年夏天，辛巴多在斯坦福大学心理学楼的地下室中建造了一个模拟监狱，用于一项由海军研究办公室资助的研究，该研究关注囚犯和军队狱警之间冲突的原因。在这里，来自全国各地的21名大学生扮演囚犯或狱警的角色，进行了为期六天的实验。然而，与1961年米尔格拉姆进行实验时不同，20世纪70年代已经有了伦理审查委员会(IRB)的存在。实际上，辛巴多和他的研究生们获得了IRB的批准。辛巴多的著作《魔鬼效应》描述了他采取的措施，以避免对受试者造成伤害，包括禁止使用身体暴力。所有参与者都经过知情同意，其中包括他们可以随时退出实验的信息。但是，与米尔格拉姆的研究不同，这项研究没有涉及欺骗。然而，研究人员给予了狱警一些建议，告诉他们应该如何行为，而不是让所有行为自然发展。这项研究被认为是有史以来最有害的研究之一。例如，它使用心理学技术引发了无聊感，剥夺了受试者的隐私、个体化和权力感。受试者对研究的反应非常强烈，扮演狱警和囚犯的学生以惊人的程度内化了他们的新身份，因此实验在计划的两周而不是六天后被终止，因为出现了心理和情绪创伤。有些囚犯情绪失控，尖叫不止，进行绝食抗议，而一些狱警表现出虐待行为，甚至用灭火器袭击囚犯。为了预防问题，研究团队在实验结束后对每个受试者进行了详细的解释，并对每个受试者进行了一年的跟踪观察以寻找不良影响。作为对批评的回应，辛巴多指出，美国心理学协会唯一收到的调查该研究的请求来自他本人，并且该研究得到了完全的批准。一项后续研究的提案被IRB拒绝，辛巴多在2007年的书中向参与者道歉。

监狱研究是一项实验室实验，通常受到高度人为因素的批评，因此辛巴多和他的同事们竭尽全力使其尽可能真实，就像米尔格拉姆一样。例如，他们说服了帕洛奥尔执法部门让真正的警察“逮捕”受试者。他们给狱警穿上

与真正监狱制服类似的服装。囚犯整天呆在有钢铁栏杆的牢房中，

iv研究人员必须富有创意，才能让警方合作。根据辛巴多在《魔鬼效应》中的描述，他们通过向警方承诺将此实验作为电视台晚间新闻的一则好宣传，来诱使警方合作。

同时，监狱看守们按照常规班次工作。他们安排了真实的亲友探访日，并进行了模拟假释听证会。151 这项研究被誉为至今为止最具真实感的典范。152 研究人员指出，尽管他们希望实验尽可能真实，但在实际监狱中进行实验并不可行，因为“现实世界或社会科学家所称的‘领域’中有太多无法控制的变量。这就是实验室研究的便利之处：实验者掌握主导权。”153 他们指出，已经在现有的刑事机构进行了许多研究，他们称之为自然实验。154 实验室实验、自然实验和野外实验之间的区别将在第4章中探讨。辛巴多解释说：“社会学家和犯罪学家已经对实际监狱生活进行了研究，但它们存在一些严重的缺点。这些研究人员无法自由观察监狱生活的所有方面......他们只能看到他们被允许看到的。”155 他们解释选择进行尽可能真实的实验室实验，是因为他们的研究问题无法使用真实的监狱和囚犯进行研究，因为需要将监狱环境的影响与其居民的特征分离开来，即真实世界中存在的混淆变量，如第1章所述。在这种情况下，进行这项研究是因为他们认为监狱中存在的暴力和残酷行为是由囚犯的反社会个性和看守的威权特征所致。因此，他们决定选择那些没有已被认为是监狱环境的原因而具有反社会和虐待个性的人（即200人），这个模拟监狱“完全由在所有基本维度上与社会其他人没有区别的个体组成。”156 因此，通过保持被认为对监狱环境负责的情境因素不变，该研究表明监狱中的行为可靠地归因于情境，而不仅仅是其中人的个性。

为了获得“正常”的受试者，研究团队结合了米尔格拉姆基于特定标准的有目的选择和费希尔的随机分配。他们通过报纸广告招募受试者，承诺每天参与可获得15美元的报酬。与米尔格拉姆不同，他认为大学生会对研究造成偏见，辛巴多有意选择了相似或同质化的学生。回应广告的七十五名学生不仅来自斯坦福大学，还来自全国各地的大学，碰巧在夏天在帕洛奥图度过。招募对象必须回答有关他们的家庭、身心健康和犯罪参与程度的问题，并接受研究人员的面试。研究人员选择了最稳定和成熟的受试者。与米尔格拉姆和当时的大多数其他研究一样，受试者都是男性。尽管他们寻找与美国人口相似的学生，但除一人外，其余受试者都是白人。

在经过慎重而有目的的选择程序后，所选的受试者随机分配到囚犯或看守的角色。随机分配到条件的目的是确保扮演看守的受试者在任何重要特征上与扮演囚犯的受试者没有差异。研究团队通过给受试者进行一系列心理测试来检查这一点，结果显示两组之间没有显著差异。随机分配起到了作用。159 第7章将详细介绍如何进行随机分配。

正如米尔格拉姆进行了前期研究一样，辛巴多和同事们在进行这项研究之前也进行了相关研究。这个想法起源于辛巴多一门课上的一个学生的项目。辛巴多随后进行了一项“野外实验”，在加利福尼亚州帕洛奥图和纽约州布朗克斯放置被废弃的汽车，并记录人们对其进行破坏的情况，就像拍摄隐藏摄像头一样。蓝领的布朗克斯和白领的帕洛奥图都是白人，衣着整洁，一些人还鼓励他们的孩子，他们都对这些汽车进行了破坏。通过这些前期研究，辛巴多得出结论，使人们感到匿名的环境可能促进反社会行为。161 第11章将讨论如何进行前期研究。

原始研究文章将监狱实验设计描述为“相对简单”，只有“一个处理变量，随机分配到‘看守’或‘囚犯’条件”。这种设计相对于如今典型的实验设计要简单得多，比如米尔格拉姆的设计，每个实验中只有一个条件或因素水平，但比起现今的实验设计来说要复杂得少得多。实验设计的类型将在第4章中详细讨论，而因子设计将在第6章中讨论。这项研究甚至没有一个正式的假设，研究人员解释这是由于研究的“探索性质”。他们报告了多个统计检验来支持提出的一个总体假设：“被分配到‘看守’或‘囚犯’的处理条件会在行为互动的行为测量、情绪状态和病理的情绪测量、对自我态度以及对这种新情境的应对和适应的其他指标方面产生显著不同的反应。”164 米尔格拉姆在他的初始论文中也没有提出假设。

和米尔格拉姆的实验一样，监狱研究使用了大量的测量工具，包括对受试者实际行为的视频和音频记录、研究人员的观察、使用自我报告的问卷和对受试者的面试。165 问卷包括情绪清单、人格测试、测量威权主义和马基雅维利主义的量表，以及包括诚实、有序、服从、活动、稳定、外向、男性化和共情的人格量表。166 第10章将更详细地讨论构建此类测量工具的方法。研究人员还使用了定量和定性技术的组合来进行数据三角化。例如，他们使用归纳法来对视频和音频记录以及观察的数据进行分类，然后将其转化为分数。

其中他们进行了统计检验。他们还定性地报告了研究结果，包括受试者的引用。研究结果的一个优点是，“在各种测量工具和不同观察者之间出现的关系模式的一致性”。167这篇论文的研究结果至今仍然是如何撰写实验论文中的这一部分的一个良好示例。大量的篇幅用于讨论研究中独特的发现、它们的意义以及背后的原因。它们与其他研究有关，而不会让整个部分看起来像另一个文献综述。168讨论还包括研究的局限性，包括样本量较小（N = 21），模拟监狱的条件与真实监狱相比较为简化，以及需求特征，在本例中，受试者知道他们正在被观察，并希望取悦研究者。该论文展示了研究人员如何通过他们所拥有的数据克服了其中的一些限制，例如分析受试者不知道自己被观察但他们的牢房被窃听的时刻。能够使用数据展示研究的局限性并不否定结论的能力是一种有用的技能。

最后，这项研究最重要的一点是全面而明确地回答了“为什么”问题。美国心理学会将其誉为心理研究如何应用于解决现实世界问题并为非学术人士所理解的典范，使研究与普通公民相关联。169作为对研究不道德性质的批评的回应，辛巴多报道了对社会的影响，并列举了公众对该研究表现出的兴趣，包括媒体报道、来电、来信、演讲请求以及对法律和政策的变更等等。170虽然辛巴多是在回应不道德处理的指责，但所有实验都应该对研究改善社会状况和解决现实世界问题的能力提出强有力的理由。

#### 总结

过去这些著名的实验有一些共同之处，其中包括创造力。这些社会科学家愿意不遗余力地确保他们的实验找到真实的东西，而不仅仅是支持一个假设。菲利普·辛巴多亲自请来了真实的警察协助他的实验。B·F·斯金纳发明了自己的测量设备。在他们的探索过程中，一些人忽视了保护被试者的责任，从而使今天的研究人员更好地理解了伤害与真实之间的权衡。实验者仍然应该努力使他们的研究条件尽可能逼近现实；例如，如果一项研究将测试电视信息的影响，那么被试者应该通过电视观看，而不是以书面脚本的形式阅读。但是，防止伤害是至关重要的。第11章将提供更具体的建议，以避免或减少实验中的欺骗行为，以及其他伦理问题。

1. 在他对坏血病的研究中，詹姆斯·林德使用了以下方式将受试者分配到治疗条件中：

a. 随机分配

b. 匹配策略

c. 代表性策略

d. 雪球策略

**测试你的知识**

• 不使研究具有现实性

• 在可以避免的情况下使用欺骗

• 复制他人的研究结果而不是创造新知识Not making studies realistic

**常见错误**

这些科学家并不害怕尝试新事物。重要的知识进步是通过知道何时借鉴他人的经验，以及何时独立开展研究而取得的。太多人害怕偏离已经做过的事情，或者以与已有测量方法不同的方式进行测量。这导致了研究只是确认他人研究结果而没有其他进展，而科学的目的是创造新知识。

这些历史上的例子还表明研究是混乱的。它并不总是像期刊上的文章所展示的那样整洁。大多数研究者在他们的著作中记录了这个混乱的过程，而不是在期刊文章中。斯金纳、米尔格拉姆和辛巴多的著作是很好的例子。

最后，这些研究显示重要的是不要伤害受试者，并尽量避免欺骗，同时遵循伦理指南（参见有关欺骗的2.3小节）。在平衡保护受试者的责任的同时，寻找有效的答案至关重要。如果可能的话，应避免使用欺骗，并在不可避免时尽量减少使用。在研究结束后向参与者进行说明，并提供研究结果的报告有助于以尊严和尊重对待人们。这些步骤还将有助于确保自己的研究因正确的原因而载入史册。

1. Designing Experiments for the Social Sciences

2. 据某些说法，最早使用随机分配的实验家之一是：

a. 詹姆斯·林德

b. 艾萨克·牛顿

c. C. S. 皮尔斯

d. 阿尔伯特·爱因斯坦

3. 罗纳德·费雪对于一位女士无法分辨茶和牛奶的先后顺序的预测是一个例子，这是一种：

a. 另一种假设

b. 零假设

c. 显而易见的假设

d. 理论假设

4. B. F. 斯金纳以使用：

a. 小样本

b. 大样本

c. 随机样本

d. 目的样本

5. 当B. F. 斯金纳的装置卡住无法工作时，他将其认定为：

a. 需要纠正的问题

b. 科学的混乱和机缘巧合的作用

c. 破坏行为

d. 控制混杂因素的方法

6. 斯坦利·米尔格拉姆在他的服从研究中使用了延迟测量、自我报告以及受试者的行为。这说明了：

a. 测量

b. 反应选择

c. 控制混杂因素

d. 最大化变异

7. 斯坦利·米尔格拉姆还使用了态度量表。这说明了：

a. 测量

b. 反应选择

c. 控制混杂因素

d. 最大化变异

Chapter 2 ■ Ethics and Famous Experiments in History **47**

8. 预备研究的目的是：

a. 发现操作是否按预期工作

b. 发现受试者是否认为实验是真实的

c. 通过试运行实验来确定需要进行哪些改变

d. 确保等效组

9. 在以下哪项研究中存在欺骗行为？

a . C. S. 皮尔斯关于重物的研究

b. 罗纳德·费雪的品茶实验

c. 斯坦利·米尔格拉姆的服从权威实验

d. 菲利普·辛巴达的监狱实验

10. 在监狱实验中，菲利普·辛巴达使用了：

a. 目的抽样

b. 随机分配

c. 既有目的抽样又有随机分配

d. 以上都不是

**Answers**

1. b

2. c

3. b

4. a

5. b

6. a

7. b

8. c

9. c

10. c



**48** Designing Experiments for the Social Sciences

• 《实验者》（The Experimenter），关于斯坦利·米尔格拉姆著名的电击实验，2015年制作。可在Netflix、Amazon和Microsoft Movies and TV上观看。该电影在2015年的圣丹斯电影节上放映。米尔格拉姆本人也拍摄了实际的实验过程，你可以在YouTube和其他视频服务上搜索“Stanley Milgram experiment video”以观看真实的录像片段，其中包括米尔格拉姆本人。

• 《斯坦福监狱实验》（The Stanford Prison Experiment），关于菲利普·辛巴达的实验，2015年制作。可在Amazon、Microsoft Movies and TV以及iTunes上观看。辛巴达拥有一个网站，上面提供了更多关于这项研究的信息：http://www.prisonexp.org。

###### 推荐阅读

没有什么比阅读原始研究更具启发性，而不仅仅是听别人说。因此，我推荐以下内容：

《模拟监狱中的人际动力学》

《群体压力与对个体的行为》

《群体压力的解放效应》

《服从与违抗权威的某些条件》

《关于感觉的微小差异》

《鸽子的迷信行为》

对于有关研究创新的利与弊的深入讨论，可以阅读：

• Voosen, Paul. 2015. "For Researchers, Risk Is a Vanishing Luxury." Chronicle of Higher Education. www.Chronicle.com. 这是一篇付费文章，意味着需要付费或订阅。请查看你所在的大学图书馆是否可以获取。

对于那些特别对欺骗历史感兴趣的人，可以阅读以下书籍：

《现实的幻觉：社会心理学中的欺骗历史》

对于一些有趣的阅读，展示这些实验家的人性面，可以阅读：

《开启斯金纳的盒子：20世纪伟大的心理实验》

Chapter 2 ■ Ethics and Famous Experiments in History **49**

g h j g k h

1. Nancy Felipe Russo和Agnes N. O'Connell，“我们过去的模型：心理学的前辈，”《女性心理学季刊》5，第1期（1980）：11–54。

2. 同上。

3. Kendra Cherry，“改变心理学的10位女性：深入了解心理学中的女性，”VeryWellMind，https://www.verywellmind.com/women-who-changed-psychology-2795260。访问日期：2018年3月21日。

4. 要了解更多心理学中女性的传记，请参阅《女性心理学季刊》1980年，第5卷，第1期的专题。

5. Agnes N. O'Connell和Nancy Felipe Russo，“成就的模型：杰出的女性心理学家，”《女性心理学季刊》5，第1期（1980）：6–10。

6. Laurel Furumoto，“Mary Whiton Calkins（1863–1930），”《女性心理学季刊》5，第1期（1980）：55–68。

7. 同上。

8. 同上。

9. Joseph Jastrow，“一项心理统计学研究，”《新评论》5（1981）：564。

10. Mary Whiton Calkins，“男女的思想共同体，”《心理评论》3（1896）：426–430。

11. Cherry，“改变心理学的10位女性，”同上。

12. “心理学家Mamie Phipps Clark简介：对少数族裔的自我概念有重要贡献的讨论，”VeryWellMind，https://www.verywellmind.com/mamie-phipps-clark-biography-2796022。访问日期：2018年3月21日。

13. L. Nyman，“记录历史：与Kenneth Bancroft Clark的采访，”《心理学史》13，第1期（2010）：74–88。

14. “心理学家Mamie Phipps Clark简介：对少数族裔的自我概念有重要贡献的讨论，”同上。

15. Stephen N. Butler，“Mamie Katherine Phipps Clark（1917–1983），”《阿肯色州历史与文化百科全书》，http://www.encyclopediaofarkansas.net/ encyclopedia/entry-detail.aspx?entryID=2938（2016）。

16. 同上。

17. Cherry，“改变心理学的10位女性，”同上。

18. M. D. Ainsworth和B. A. Wittig，“一岁婴儿在陌生情境中的依恋和探索行为，”收录于《婴儿行为的决定因素》，编者B. M. Foss（伦敦：Methuen，1969），113–136页。

19. U. G. Wartner等，“南德国的依恋模式，”《儿童发展》65（1994）：1014–1127页。

20. M. E. Lamb，“生命的第二年中母婴和父婴依恋的发展，”《发展心理学》13，第6期（1977）：637–648页。

21. M. Marrone，《依恋与互动》，（纽约：Jessica Kingsley，1998）；E. C. Melhuish，“爱的度量？对依恋评估的概述，”《ACPP评论与通讯》15（1993）：269–275页。

22. Russo和O'Connell，“我们过去的模型，”同上。

23. 美国国家医学图书馆，“坏血病，”Medline Plus（2016年）。

24. James Lind，《坏血病论文：分三部分。对该病的性质、原因和治疗的研究。以及对该主题已经发表的论述的批判和历史的观点》，（爱丁堡：Sands, Murray, and Cochran出版，由A. Millar出版，1753年）。

25. 同上，191–192页。

26. 同上，193页。

27. 同上。

28. R. Barker Bausell，《进行有意义的实验：成为科学家的40个步骤》，（加利福尼亚州千橡市：Sage，1994年）。

29. Lind，《坏血病论文，》同上。

30. Stephen M. Stigler，“早期国家的数理统计学，”《统计学年鉴》6，第2期（1978）：239–265页。

31. Charles Sanders Peirce和Joseph Jastrow，“关于感觉中的小差异，”《国家科学院院刊》3（1885）：75–83页。

32. Stephen M. Stigler，“心理学和教育研究中统计概念的历史观点，”《美国教育杂志》101，第1期（1992年11月）：60–70页。

33.查尔斯·桑德斯·皮尔斯（Charles Sanders Peirce）和约瑟夫·贾斯特罗（Joseph Jastrow）的文章《关于感觉中的微小差异》可在《19世纪美国数学统计的贡献》（American Contributions to Mathematical Statistics in the Nineteenth Century），史蒂芬·M. 斯蒂格勒（Stephen M. Stigler）编辑，纽约：Arno Press，1980/1885，第80页找到。

34. 查尔斯·桑德斯·皮尔斯（Charles Sanders Peirce），《科学哲学论文集》（Essays in the Philosophy of Science），纽约：Liberal Arts Press，1957，第217页。

35. 同上，第104页。

36. Joseph Brent，《查尔斯·桑德斯·皮尔斯：一生》（Charles Sanders Peirce: A Life），第二版，伊迪安纳波利斯：印第安纳大学出版社，1998年。

37. Stigler，《19世纪早期的数学统计》（Mathematical Statistics in the Early States），第249页。

**50** Designing Experiments for the Social Sciences

38. David Salsburg，《品茶的女士：统计如何在20世纪革命科学》（The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century），纽约：W. H. Freeman，2001年。

39. Ronald A. Fisher，《研究工作者的统计方法》（Statistical Methods for Research Workers），爱丁堡：Oliver and Boyd，1925年。

40. Diana C. Mutz和Robin Pemantle，《实验研究的标准：鼓励更好地理解实验方法》（Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods），《实验政治学杂志》（Journal of Experimental Political Science）2卷2期（2016年）：192-215页。

41. Salsburg，《品茶的女士》（The Lady Tasting Tea）。

42. 同上。

43. Stigler，《统计学的历史观》（A Historical View of Statistical）。

44. Salsburg，《品茶的女士》。

45. Ronald A. Fisher，《实验设计》（The Design of Experiments），爱丁堡：Oliver and Boyd，1937年。

46. Salsburg，《品茶的女士》，第8页。

47. Robert P. Hawkins，《Burrhus Frederick Skinner的生平和贡献》（The Life and Contributions of Burrhus Frederick Skinner），《教育与儿童治疗》（Education & Treatment of Children）13卷3期（1990年）：258页。

48. B. F. Skinner，《科学方法的案例历史》（A Case History in Scientific Method），《美国心理学家》（The American Psychologist）11卷5期（1956年）：221-233页。

49. Joshua B. Plavnick和Summer J. Ferreri，《教育研究中的单一案例实验设计：教学与学习中的因果分析方法》（Single-Case Experimental Designs in Educational Research: A Methodology for Causal Analyses in Teaching and Learning），《教育心理学评论》（Education Psychology Review）25卷（2013年）：549-569页；Denise A. Soares等，《当代课堂环境中代币经济使用的效应大小：单一案例研究的荟萃分析》（Effect Size for Token Economy Use in Contemporary Classroom Settings: A Meta-Analysis of Single-Case Research），《学校心理学评论》（School Psychology Review）45卷4期（2016年）：379-399页；Stephen F. Ledoux，《行为主义百年回顾》（Behaviorism at 100），《美国科学家》（American Scientist）100卷1期（2012年）：60-65页；Walter R. Nord，《超越教学机：管理理论与实践中被忽视的操作条件》（Beyond the Teaching Machine: The Neglected Area of Operant Conditioning in the Theory and Practice of Management），《组织行为与人类绩效》（Organizational Behavior & Human Performance）4卷4期（1969年）：375-401页；Laura Vandeweghe等，《促进青少年健康饮食的认为有效和可行策略：系统回顾和荟萃分析》（Perceived Effective and Feasible Strategies to Promote Healthy Eating in Young）"儿童：与父母、家庭儿童保育提供者和托儿所助理的焦点小组讨论"（Children: Focus Groups with Parents, Family Child Care Providers and Daycare Assistants），发表于《BMC公共卫生》（BMC Public Health）16卷（2016年）：1-12页。

50. Ledoux, "行为主义百年回顾"（Behaviorism at 100）。

51. Skinner, "科学方法的案例历史"（A Case History in Scientific Method），第227页。

52. Skinner，《生物体的行为：一项实验分析》（The Behavior of Organisms: An Experimental Analysis），纽约：Appleton-Century，1938年。

53. Skinner, "科学方法的案例历史"。

54. Skinner，《生物体的行为：一项实验分析》。

55. Skinner, "科学方法的案例历史"。

56. Alexandra Rutherford, "从实验室到生活的B. F. Skinner"（B. F. Skinner From Laboratory to Life），收录于《战后社会科学史研讨会系列》（History of Postwar Social Science Seminar Series），伦敦，2007年。

57. Alexandra Rutherford, "B. F. Skinner在美国生活中的行为技术：从消费文化到反文化"（B. F. Skinner's Technology of Behavior in American Life: From Consumer Culture to Counterculture），《行为科学史杂志》（Journal of the History of the Behavioral Sciences）39卷1期（2003年冬季）：1-23页。

58. Plavnick和Ferreri，《单一案例实验设计》（Single-Case Experimental Designs）；T. R. Kratochwill等，《单一案例设计技术文档》（Single-Case Designs Technical Documentation）（2010年）。http://ies.ed.gov/ncee/wwc/Document/229。

59. Plavnick和Ferreri，《单一案例实验设计》；Hawkins，《Burrhus Frederick Skinner的生平和贡献》（The Life and Contributions of Burrhus Frederick Skinner）。

60. Skinner, "科学方法的案例历史"。

61. 同上，第227页。

62. B. F. Skinner，《鸽子的迷信行为》（Superstition in the Pigeon），《实验心理学杂志》（Journal of Experimental Psychology）38卷（1948年）：168-172页。

63. Skinner, "科学方法的案例历史"，第229页。

64. 同上，第223页。

Chapter 2 ■ Ethics and Famous Experiments in History **51**

65. Hawkins，《Burrhus Frederick Skinner的生平和贡献》。

66. Charles B. Ferster和B. F. Skinner，《强化计划》（Schedules of Reinforcement），纽约：Appleton-Century-Crofts，1957年。

67. Skinner, "科学方法的案例历史"。

68. 同上。

69. 同上，第224页。

70. 同上，第224页。

71. D. E. Blackman和R. Pellon，《B. F. Skinner对行为药理学跨学科科学的贡献》（The Contributions of B. F. Skinner to the Interdisciplinary Science of Behavioural Pharmacology），《英国心理学杂志》（British Journal of Psychology）84卷1期（1993年）：1页。

72. Skinner, "科学方法的案例历史"，第226页。

73. 同上，第224页。

74. 同上，第225页。

75. 同上，第232页。

76. Skinner，《生物体的行为：一项实验分析》。

77. Skinner，《科学方法的案例历史》，第223页。

78. Elisha Babad, Eyal Peer和Yehonatan Benayoun，《在单一情境中是否可能存在多种偏见？来自媒体偏见研究的证据》（Can Multiple Biases Occur in a Single Situation? Evidence from Media Bias Research），《应用社会心理学杂志》（Journal of Applied Social Psychology）42卷6期（2012年）：1486-1504页。

79. 同上，第1501-1502页。

80. Diana Baumrind，《关于研究伦理的一些思考：阅读Milgram的《服从的行为研究》之后》（Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience'），《美国心理学家》（American Psychologist）19卷6期（1964年）：421-423页。

81. Stanley Milgram，《服从权威：一种实验观点》（Obedience to Authority: An Experimental View），纽约：Harper & Row，1974年。

82. Milgram，《服从和不服从权威的一些条件》（Some Conditions of Obedience and Disobedience to Authority），《人际关系》（Human Relations）18卷1期（1965年）：57-76页。

83. Milgram，《服从权威：一种实验观点》。

84. Diana Baumrind，《使用有意的欺骗进行研究：伦理问题的再探讨》（Research Using Intentional Deception: Ethical Issues Revisited），《美国心理学家》（American Psychologist）40卷2期（1985年）：165-174页。

85. Davide Barrera和Brent Simpson，《关于欺骗的许多麻烦：社会科学研究中欺骗研究参与者的后果》（Much Ado About Deception: Consequences of Deceiving Research Participants in the Social Sciences），《社会学方法与研究》（Sociological Methods & Research）41卷3期（2012年）：383-413页。

86. 同上；D. Geller，《替代欺骗：为什么、什么以及如何？》（Alternative to Deception: Why, What, and How?），收录于《社会研究的伦理学：调查和实验》（The Ethics of Social Research: Surveys and Experiments），Joan E. Sieber编辑（纽约：斯普林格出版公司，1982年），第38-55页。

87. Barrera和Simpson，《关于欺骗的许多麻烦》。

88. Karen A. Hegtvedt，《伦理和实验》（Ethics and Experiments），收录于《社会科学中的实验室实验》（Laboratory Experiments in the Social Sciences），Murray Webster和Jane Sell编辑（伦敦：爱尔维尔，2014年）：23-51页。

89. 同上。

90. Baumrind，《使用有意的欺骗进行研究》。

91. Immo Fritsche和Volker Linneweber，《非反应性（潜在）方法》（Nonreactive (Unobtrusive) Methods），收录于《心理测量手册：多方法视角》（Handbook of Psychological Measurement—A Multimethod Perspective），M. Eid和E. Diener编辑（华盛顿，DC：美国心理学协会，2004年）。

92. 同上。

93. Baumrind，《使用有意的欺骗进行研究》（Research Using Intentional Deception），第166页。

**52** Designing Experiments for the Social Sciences

94. Philip Zimbardo，《关于人类心理研究中干预伦理：以斯坦福监狱实验为例》（On the Ethics of Intervention in Human Psychological Research: With Special Reference to the Stanford Prison Experiment），《认知》（Cognition）2卷2期（1973年）：243-256页。

95. Philip G. Zimbardo, S. M. Andersen和I. G. Kabat，《诱发听力缺陷引发实验性偏执狂》（Induced Hearing Deficit Generates Experimental Paranoia），《科学》（Science）212卷（1981年6月）：1529-1531页。

96. G. D. Marshall和P. G. Zimbardo，《不充分解释的生理激动引发的情感后果》（Affective Consequences of Inadequately Explained Physiological Arousal），《个性与社会心理学杂志》（Journal of Personality and Social Psychology）37卷（1979年）：970-988页。

97. M. E. Shanab和K. A. Yahya，《儿童服从行为的行为研究》（A Behavioral Study of Obedience in Children），《个性与社会心理学杂志》（Journal of Personality and Social Psychology）35卷（1977年）：530-536页。

98. 同上，第534页。

99. 同上，第536页。

100. Stanley Milgram，《群体压力的解放效应》（Liberating Effects of Group Pressure），《个性与社会心理学杂志》（Journal of Personality and Social Psychology）1卷（1965年）：127-134页。

101. Sandra D. Nicks、James H. Korn和Tina Mainieri，《社会心理学和人格研究中欺骗的兴衰，1921年至1994年》（The Rise and Fall of Deception in Social Psychology and Personality Research, 1921 to 1994），《伦理与行为》（Ethics & Behavior）7卷1期（1997年）：69页。

102. 同上。

103. 同上。

104. 同上。

105. J. R. McNamara和K. M. Woods，《心理研究中的伦理考虑：一项比较综述》（Ethical Considerations in Psychological Research: A Comparative Review），《行为疗法》（Behavior Therapy）8卷（1977年）：703-708页。

106. J. Seeman，《心理研究中的欺骗》（Deception in Psychological Research），《美国心理学家》（American Psychologist）24卷（1969年）：1025-1028页。

107. Ralph Hertwig和Andreas Ortmann，《经济学实验中的实验实践：对心理学家的方法论挑战？》（Experimental Practices in Economics: A Methodological Challenge for Psychologists?），《行为与脑科学》（Behavioral & Brain Sciences）24卷3期（2001年）：383页；R. Hertwig和A. Ortmann，《实验中的欺骗：重审对其辩护的论点》（Deception in Experiments: Revisiting the Arguments in Its Defense），《伦理与行为》（Ethics & Behavior）18卷（2008年）：59-92页。

108. Hegtvedt，《伦理与实验》（Ethics and Experiments）。

109. James H. Korn，《现实的幻觉：社会心理学中欺骗的历史》（Illusions of Reality: A History of Deception in Social Psychology），阿尔巴尼，纽约：纽约州立大学出版社，1997年。

110. Baumrind，《使用有意的欺骗进行研究》（Research Using Intentional Deception）。

111. Hegtvedt，《伦理与实验》。

112. Baumrind，《使用有意的欺骗进行研究》。

113. 同上。

114. 同上。

115. 同上。

116. Barrera和Simpson，《欺骗问题的大惊小怪》（Much Ado About Deception）；S. Bonetti，《实验经济学和欺骗》（Experimental Economics and Deception），《经济心理学杂志》（Journal of Economic Psychology）19卷3期（1998年）：377-395页。

117. Hegtvedt，《伦理与实验》。

118. 同上。

119. Baumrind，《使用有意的欺骗进行研究》。

120. Joan E. Sieber、Rebecca Iannuzzo和Beverly Rodriguez，《心理学中的欺骗方法：23年来是否有变化？》（Deception Methods in Psychology: Have They Changed in 23 Years?），《伦理与行为》（Ethics & Behavior）5卷1期（1995年）：67页。

121. Baumrind，《使用有意的欺骗进行研究》；S. A. McLeod，《心理学研究伦理》（Psychology Research Ethics），Simply Psychology http://www.simplypsychology.org/Ethics.html。

122. Baumrind，《使用有意的欺骗进行研究》；McLeod，《心理学研究伦理》。

123. Sieber、Iannuzzo和Rodriguez，《心理学中的欺骗方法》，83页。

124. Milgram，《服从权威：一种实验观点》（Obedience to Authority: An Experimental View），23页。

125. Milgram，《服从的行为研究》（Behavioral Study of Obedience），《异常与社会心理学杂志》（Journal of Abnormal and Social Psychology）67卷4期（1963年）：371-378页。

126. 同上；Milgram，《群体压力和对个人的行动》（Group Pressure and Action Against a Person），《异常与社会心理学杂志》（Journal of Abnormal and Social Psychology）69卷（1964年）：137-143页。

Chapter 2 ■ Ethics and Famous Experiments in History **53**

134. 同上，92页。

135. Milgram，《群体压力和对个人的行动》（Group Pressure and Action against a Person）。

136. Milgram，《服从的行为研究》（Behavioral Study of Obedience），373页。

137. 同上，374页。

138. Milgram，《服从权威：一种实验观点》（Obedience to Authority: An Experimental View），203页。

139. 同上，205页。

140. 同上，206页。

141. 同上，22页。

142. 同上，22页。

143. Rebecca S. McEntee、Renita Coleman和Carolyn Yaschur，《比较生动写作和照片对公共关系中道德判断的影响》（Comparing the Effects of Vivid Writing and Photographs on Moral Judgment in Public Relations），《新闻与大众传播季刊》（Journalism and Mass Communication Quarterly）94卷4期（2017年）：1011-1030页。

144. Philip Zimbardo，《路西法效应：理解善良的人如何变坏》（The Lucifer Effect: Understanding How Good People Turn Evil），纽约：Random House，2007年。

145. 同上，508页。

146. 同上。

147. Craig Haney、Curtis Banks和Philip Zimbardo，《模拟监狱中的人际动力》（Interpersonal Dynamics in a Simulated Prison），《犯罪学与刑罚学国际期刊》（International Journal of Criminology and Penology）1卷（1973年）：69-97页；Zimbardo，《路西法效应》（The Lucifer Effect）。

148. Haney、Banks和Zimbardo，《模拟监狱中的人际动力》；Zimbardo，《论人类心理学研究中的干预伦理》（On the Ethics of Intervention in Human Psychological Research）；Zimbardo，《路西法效应》。

149. Zimbardo，《论人类心理学研究中的干预伦理》。

150. Zimbardo，《路西法效应》。

151. 同上。

152. Murray Webster和Jane Sell，《社会科学实验室实验》（Laboratory Experiments in the Social Sciences），阿姆斯特丹：Elsevier，2007年。

153. Zimbardo，《路西法效应》，36页。

154. Haney、Banks和Zimbardo，《模拟监狱中的人际动力》，89页。

155. Zimbardo，《路西法效应》，32页。

156. Haney、Banks和Zimbardo，《模拟监狱中的人际动力》。

157. 同上，21页。

158. 同上，21页。

159. 同上。

160. 同上，24页。

161. Zimbardo，《路西法效应》，24页。

162. Haney、Banks和Zimbardo，《模拟监狱中的人际动力》，73页。

163. 同上，77页。

164. 同上，72页。

165. 同上，69页。

166. 同上，73页。

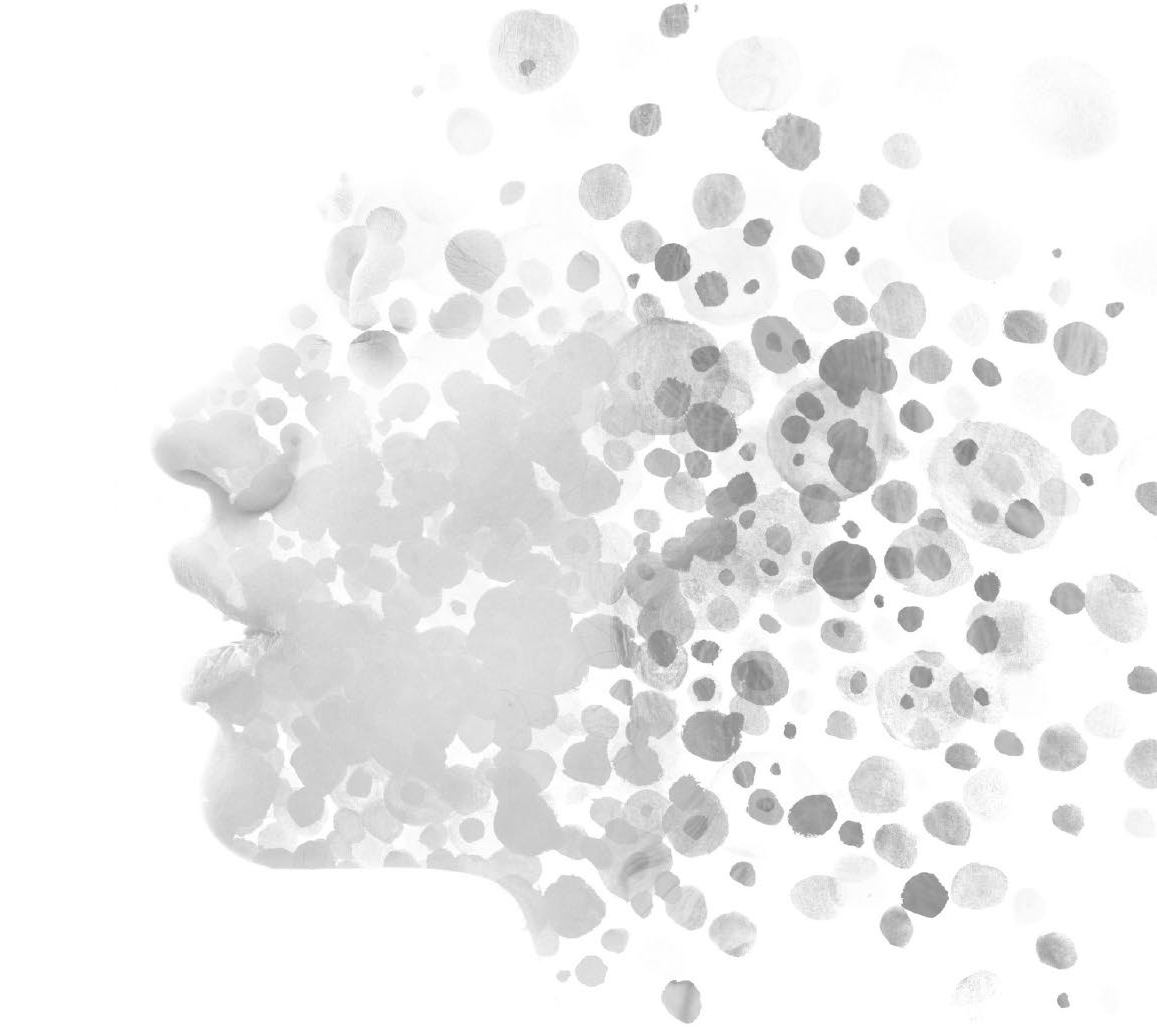
167. 同上，78页。

168. 同上。

169. George A. Miller，《在80年代将心理学传播出去》（Giving Psychology Away in the '80s），《今日心理学》（Psychology Today）13期（1980年）：38页及以下。

170. Zimbardo，《论人类心理学研究中的干预伦理》（On the Ethics of Intervention in Human Psychological Research）。

171. 关于“那又怎样”问题的扩展陈述，请参阅《路西法效应》前言，第x至第xii页。



3

# 理论、文献和假设

*完全成功的干预措施从未仅仅通过头脑风暴而出现。它们总是基于理论、先前的研究或丰富的临床经验提出的。 ——R. Barker BausellFull-blown*

• 解释理论和文献在实验中的作用。

• 准备一份文献综述，将先前的工作与你自己的研究联系起来，并做出理论贡献。

• 描述假设和研究问题在实验中的作用。

• 总结何时预测假设中的差异或方向。

• 使用给定的图表制定原创假设。

学习目标

根据Bausell的说法，我们应该补充一点，即实验并非用于探索性研究，即在问题处于初步阶段时使用。相反，实验需要相当丰富的理论和先前研究的证据。因此，在开始任何实验之前，进行充分的准备工作非常重要，其中包括对文献进行彻底的搜索，包括理论方面的文献。



iStock.com/baramee2554

本章在前一章的基础上，通过考察前人所发现的角色，或者说像艾萨克·牛顿所说的“站在巨人的肩膀上”的概念，进一步探讨了这一角色的重要性。尽管本章的内容也适用于其他研究方法，但实验特别强调通过建立在先前发现的基础上来创造新知识。本章还涵盖了“良好执行的实验的七个属性”中的两个，包括理论解释或对正在测试的理论进行解释，以及明确确定研究问题和假设。本章从如何解释正在测试的理论，将其与实验中使用的变量联系起来，并将所有假设与该理论联系起来的角度来审视论文的文献综述部分。假设在实验中是特别强大的分析工具，因此，了解如何撰写文献综述以及由此产生的假设对于这种方法尤为重要。

正如第1章所介绍的，设计实验的第一步是明确要测试的因果关系，并在目的陈述中清晰表达出来。但这些关系不应仅基于直觉或好奇心。最好是，要测试的想法应该建立在对领域和其中的文献和理论的理解之上。衡量因果关系需要理论思考；否则，实验将成为一个凭运气的试错过程。(关于理论的更多信息，请参见3.1更多关于理论的内容框。)

社会科学知识发展的一种常见途径是通过使用观察、访谈、人类学研究等方法进行探索性工作。这些类型的定性方法明确设计用于描述和生成理论。它们揭示了潜在的含义——即在表面上不容易看出的含义——或者帮助我们理解现象的本质并预测其发生的条件。但由于其固有特性(样本小、非随机抽取、缺乏控制等)，它们不适合测试或完善生成的理论。这就是实验的主要目的之一。在通过定性或其他方式生成理论之后，研究可以进一步采用测试它们的方法。调查可以很好地获取关于行为的意见、态度和自我报告。内容分析可以揭示各种类型的消息、文本或视觉特征。虽然这些方法产生的研究增加了新的知识并扩展了理论，但它们也提供了可能的原因，并提供了相关性证据，为实验工作奠定了基础。随后，实验研究能够确定在自然环境中观察到的现象是否可以在受控环境中重复出现，

For Review-No Commercial Use(2023)

### MORE ABOUT . . . BOX 3.1

Theory

理论是对某一特定现象的概括性解释，或者说是一个“组织起来的概念和原理体系”。这种学术说法实际上是指理论是关于现实世界中实际发生的事情的。引用门禁机制之父库尔特·莱温的话来说，“没有比一个好的理论更实用的了。”虽然理论可能看起来神秘而难以理解，但它实际上只是一种理解生活中发生的事情的工具。当持续支持假设时，理论会不断发展。一个好的理论不仅应该描述现象，还应该解释现象并能够对其进行预测。例如，框架理论的一个方面描述了新闻报道如何构建为离散的事件——比如一起犯罪案件的报道——或更普遍地作为增加背景的主题——比如关于犯罪的报道，而不仅仅是某一个具体案例。它解释了人们如何感知这些类型的框架，并且还可以预测人们对其的反应——例如，在事件以离散的方式被框架时，将个人视为责任人；而在事件以主题方式被框架时，将责任归咎于社会因素。实验通过测试假设和扩展我们现有的理解来建立理论。

然而，在生物学和相关学科中，有一些关于“探索性实验”的讨论。这些研究人员所指的探索性实验并不是无目标、无理论基础的数据收集项目。相反，探索性实验是指在理论指导或基础下进行的实验，但不一定通过正式的假设来测试它。这些研究具有理论背景，而不是完全缺乏理论框架。理论仍然“对探索性实验的成功至关重要”。以“探索性”为借口来证明缺乏理论、功效不足、受试者非随机分配、未控制变量或其他弱点是不合适的。在这些情况下，研究人员应将其视为试点研究，吸取经验教训，并进行改进以消除弱点，而不是试图将其作为探索性研究发表。

在实验中，需要排除所有可能干扰结果或对现实世界中的可行替代解释负责的许多变量。当确定实验是下一个合乎逻辑的方法时，第一步是进行彻底的文献和特定理论的搜索。

这并不是说研究人员永远不能通过其他方式提出实验的想法，比如观察并对为什么发生特定现象产生好奇，就像詹姆斯·林德在他对坏血病的研究中所做的那样。通过这种方式产生的想法当然可以通过实验进行测试。但是，进行测试时应该有一些基于理论或证据的原因。在追寻由研究问题产生的时候，理论或证据应该为测试这些问题提供支持。

在进行观察时，找到可以帮助引导和完善它们的经验证据和理论非常重要。这一目的可以通过将两个变量联系起来的相关研究来实现，也可以通过使用不同的受试群体或不同的因变量(DV)的实验来实现。例如，在一项研究中，探讨了通过展示照片是否可以提高记者的道德判断力，以往没有使用道德判断作为因变量(DV)，照片作为自变量(IV)的研究。然而，照片已经被证明可以改善其他认知过程，如细化、移情和参与度。所有这些因变量都与道德判断相关，这是其他研究所证明的。这些理论和经验的基础为自变量和结果变量提供了联系，虽然间接。这个想法源于对于当涉及到照片时记者的伦理推理如何改善的文章和其他非经验性作品的大量猜测。

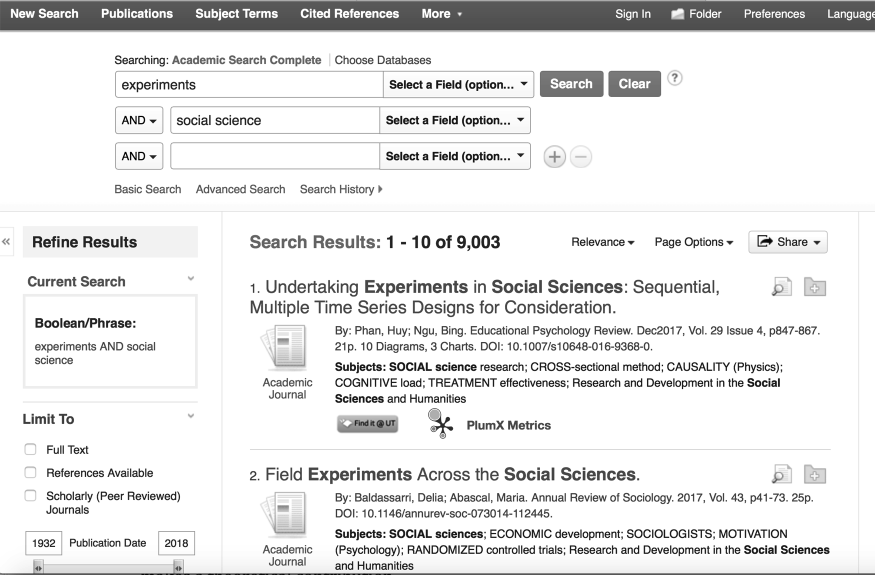
#### 文献综述

For Review-No Commercial Use(2023)

在论文的文献综述部分，可以对使用的理论及其与研究变量之间的关系进行解释。一篇完成良好的文献综述为整个实验提供了坚实的基础，然而，Bausell指出，“大多数初学者最常违反的单一规则”是对相关文献了解不够充分。有经验的研究人员可能不觉得有必要在进行实验之前撰写正式的文献综述，因为他们始终关注最新的研究发现并熟悉他们所了解的理论。但是对于其他人来说，在形成任何可测试的假设之前，应该进行对相关文献的彻底搜索，并熟悉相关理论。这个过程的第一步是进行文献搜索，第二步是在论文的文献综述部分进行撰写，这也是本章的主题。

阅读他人的研究可以成为实验构想的重要源泉。首先，阅读现有的研究可以了解是否已经有人做过类似的研究，以避免只是简单地重复别人的工作。它还可以揭示已经完成的工作中存在的空白以及需要进一步开展的工作，以完善理论。每篇研究文章末尾的“局限性”和“未来研究”部分也可以提供更多的思路。这包括需要复制的研究以及对理论进行扩展和改进的想法。阅读文献还可以提供方法论的帮助，包括他人是如何测量变量的，可能需要多少被试以及新颖程序的想法。

i在某些学科领域中，不在某种相关方式上扩展理论或实证知识的复制研究可能会受到不赞同的看法。



这本书提供了如何进行文献检索和撰写文献综述的指导，以及可能已经进行和撰写了一篇或多篇文献综述的读者。由于实验在设计阶段需要投入大量工作，并且一旦收集到数据就无法更改，因此对实验的文献检索尤为重要。我建议在开始收集数据之前，实际上先起草一份文献综述，无论你对该主题有多熟悉。我发现写作的过程会减慢速度，让我有时间以一种阅读时不会有的方式思考文献。它还迫使我综合各种不同的研究结果，有助于更好地分析意义并找到可能需要填补的空白。制作一张图表可以更清晰地展示事物，有时这些图表最终会出现在发表的研究中，例如表格3.1。

在实验中，需要隔离或控制除了被操纵的变量之外可能影响结果或因变量的所有因素，因此了解所有这些因素尤为关键。然后可以对它们进行控制，或将其用作中介变量、调节变量或共变量。如果研究人员不知道它们是什么，并且未能测量或控制它们，整个实验的结果的有效性将受到威胁。

有许多关于如何进行文献检索和撰写文献综述的好书和文章，因此在这里不会详细描述它们。相反，下面是一些建议，以避免文献综述中的一些常见问题。

|  |  |  |
| --- | --- | --- |
| **表格3.1 想象力能力的指标 (Liao等人，2016年)** | | |
| **指标** | **定义** | **参考文献** |
| 新颖性 | 生成独特思想的能力 | Beaney（2005）和Vygotsky（2014/1967） |
| 生产力 | 通过广泛运用自由联想的方式产生思维的能力 | Folkm Folkmann（2010年）和Gaut（2005年） |
| 专注力 | 通过专注和沉浸来形成思维的能力 | Csikszentmihalyi (1996) 和 Folkmann (2010) |
| 敏感性 | 在创造过程中唤起情感的能力 | 瑞克尔（1978年）和谢弗勒（1986年） |
| 直觉 | 能够与目标产生直接联想的能力 | Reichling（1990年）和Townsend（2003年） |
| 有效性 | 能够生成相关而深刻的思想以实现目标 | Gilbert和Reiner（2000）以及Shin（1994） |
| 辩证法 | 通过逻辑分析可能性和替代方案寻求改进的能力 | Cartwright and Noone (2006) and Reiner and Gilbert (2000) |
| 探索能力 | 对未知事物进行探究的能力 | Coiello (2007) and Thomas (2004) |
| 具体化 | 通过使用具体的例子来可视化抽象概念的能力 | 德里克斯（1988）和维果茨基（2014/1967） |
| 转化 | The ability to perform tasks by applying information acquired across multiple fields of knowledge | Kunzendorf (1982) 库恩岑多夫  Liu and Noppe-Brandon (2009) 刘和诺普布兰登 |

来自：Liao, Kai-Hung, Chi-Chen Chang, Chao-Tung Liang, and Chaoyun Liang. 2016. “In Search of the Journalistic Imagination.”

Thinking Skills and Creativity 19: 9–20.

实验设计。这些建议并不仅适用于这种方法，但它们解决的问题在实验中经常发生，因此值得一提。

实验文献综述写作的建议

**这不是一篇书评**

重要的是分析和综合文献搜索中找到的各项研究的结果。很多时候，文献综述读起来像是书评。它们只是对之前研究结果的总结，但是没有对这些结果进行分析，展示它们之间以及与即将进行的研究之间的联系。不要假设读者对这个主题已经很熟悉；因此，需要明确说明所综述的研究之间以及与当前研究的关联。

对他们来说并不像对你一样显而易见，所以需要详细说明。此外，实验是基于先前的研究提出假设并进行测试的，因此对这种方法来说，对这些研究结果进行分析尤为重要。

与其简单地串联摘要的总结（例如，“Smith和Jones发现...”；“White、Brown、Green和Black展示了...”；“Tinker和Taylor研究了...”），一个好的文献综述应该是对这些研究的批判性分析，综合研究结果，并提出自己的新观点。这种方法也避免了看起来像是显眼地提及他人的名字，或试图提高作者的引用次数或期刊的影响因子（衡量期刊重要性的指标）的行为。

作为分析和综合的一个例子，我会讲述一个同事和我进行议程设置研究的故事，我们想要了解新闻议题或候选人特质对选民的影响哪个更为强大。我们进行了文献搜索，注意到媒体议程与公众议程之间的相关性（议程设置理论的基本前提）在特质方面要高于议题方面。这成为我们得出以下假设的基础：特质议程设置比议题议程设置更具影响力。这并非我们在文献中发现的，而是通过对理论的批判性思考和比较所有研究结果而发现的。

iStock.com/wellesenterprises

**不要成为维基百科**

许多文献综述都像维基百科条目一样。相反，一个好的综述应该只覆盖与研究相关的文献。例如，如果实验是关于二级议程设置的，可能需要简要提及一级议程设置是什么，并说明它与二级议程设置的区别，但不需要详细解释一级议程设置，总结每个相关研究，描述所有的中介变量、调节变量或条件限制情况——即它起作用的时候和不起作用的时候。相反，提及一级议程设置并说明它与本实验要测试的内容有何不同，然后继续进行。换句话说，如果研究的是B，就只关注B，即使还有A。确保涵盖所有相关的文献；有些作者在另一个方向上做得太过头了，忽略了重要的工作。关键是找到平衡点——涵盖适量的内容，不要遗漏任何真正具有重要影响的研究或经典作品。这也适用于文献综述中涉及的理论。有许多重叠的理论可以用来支持一个实验，但一个研究应该着重于最相关的理论，从而使其与实验的目的和问题紧密联系起来。综述中的理论部分应该解释为什么选择这些理论，并阐明它们如何为实验提供理论基础和指导。

### 研究聚焦 3.2

Shoemaker, P., J. Tankard, and D. Lasorsa. 2003. 《如何构建社会科学理论》. Thousand Oaks, CA: Sage.

初学者经常很难理解理论的概念，特别是如何构建理论。期刊审稿人常常会问作者：“你的理论贡献是什么？”、“这项研究提供了哪些理论观点？”以及“这项研究在理论构建方面有何贡献？”找到回答这些问题的方法常常让许多新兴社会科学研究人员彻夜难眠。当这本关于如何在社会科学中构建理论的书面世时，它以其逻辑严谨、循序渐进的方法直接吸引了我作为一个内心深处的实验者。它兑现了它教授“以最小困难进行理论构建这一挑战性活动”的承诺。这本书首先向读者保证他们并不孤单，承认自己对于如何构建理论和提供理论贡献的理解不足并没有什么可羞愧的。

这本书的章节逻辑清晰，首先描述了假设的书写。然后解释了理论概念，称其为“理论的构建基石”。它演示了如何将概念组合成理论陈述，广泛描述变量之间的关系。它展示了如何综合不同的部分并建立关联，无论是理论上的还是操作上的。然后它展示了如何描述和解释概念之间的关系，以提供解释力。章节涉及如何关联两个变量、三个变量、四个及更多变量，以及如何在完整理论之前构建模型。

这本书提供了社会科学学者可以理解的例子。其中一个例子是乔治·格伯纳和同事们对“培养理论”的发展，展示了它是如何起源于解释电视对观众的影响的一种方式。其中一个假设提出观看电视塑造观众的信念、意识形态和世界观。随着研究的进行和假设的支持，该理论逐渐增加了其他概念，如“共鸣”和“主流化”，指定了产生这种效应所必需的条件。进一步的研究导致理论的修订和补充，包括区分基于现实和信念的一级和二级效应的感知。

《如何构建社会科学理论》强调了创造力在构建理论中的重要性，这也是本书的一个重要观点。它提供了具体、实用的方法来为理论做出贡献。在我阅读这本书之前，我对如何发展理论从未真正理解过。

通常只使用一两个理论。例如，有至少十种理论可以用来解释通过观看电视引起的攻击行为。在使用其他研究方法的研究中，可能需要回顾许多相似的理论，而不是在旨在测试特定理论的实验中。找到最适合实验目的的理论，并为选择该理论而写出充分的理由，而不是回顾每个可能的理论。有时，两种理论会导致相反的预测，因此可以在一项研究中并列使用它们。在这个过程中，应该建立起自变量、因变量和中介变量之间的紧密联系，即治疗和结果之间的因果关系。

**进行理论贡献**

迄今为止，许多研究无论是哪种类型，都会在总结一个理论之后，然后进行作者打算进行的研究，但最终却没有回到理论这一点上。他们通常没有明确说明研究意图对该理论或多个理论做出何种贡献。对于实验研究而言，这是一个特别严重的疏忽，因为进行实验的主要原因之一就是为了发展、验证、完善或以其他方式对理论做出贡献。仅仅将理论作为基础而后弃之是不够的。必须明确指出这个实验将为哪个理论或多个理论做出何种贡献，并且在论文的讨论和结论部分对此进行扩展。（关于如何建立理论的书籍概要，请参阅研究重点 3.2。）

为实现这一目标，重要的是清晰地将所有变量（因变量、自变量、因果机制和个体差异变量）之间的关系进行链接和说明。首先，说明自变量将如何影响因变量，然后解释这将通过哪些理论过程发生。如果有助于理解，可以制作一个模型来展示这一过程，并用图表形式来可视化（例如，请参阅表格 3.1）。举个例子，前面提到的有关照片对道德判断的影响的研究感兴趣的是建立记者暴露于照片时与其道德判断水平之间的因果关系，该研究推测这是因为照片已被证明能够增加移情能力，进而加深参与感，并鼓励中心路线加工，而所有这些因素都被发现会增加道德判断。该研究的理论贡献不仅在于照片有能力提高道德判断——这是一项以前未经测试的干预措施——还在于展示了为什么照片有效的理论解释。

**将各种研究结果联系起来**

最后，一个好的文献综述的另一个重要特点是将其他研究与你自己的研究联系起来，我称之为连接各个点。在审查共同主题的每一批研究之后，明确地说明它们与即将进行的研究的关系。它将以相同的方式进行测量吗？它们是否存在需要填补的空白？这些研究中的哪些连接将有助于这个实验的开展？始终将其他研究的发现与你的研究联系起来，将其置于更大的研究领域的背景之中。读者不应该在超过一两个段落的距离内才能找到为什么要审查特定研究以及它们如何与当前研究相关联的答案。例如，在前面提到的研究中，研究了各种议程设置研究的相关性，论文解释了它们与当前研究的相关性，如下所述：“所有这些二级相关性与最高

**如何做到这一点3.3**

连接句的示例

在文献综述中，将研究与其他研究以及你自己的研究联系起来，将你的工作置于其他研究的背景中非常重要。以下是一些关于如何明确表达这种联系的研究者在他们的实验中的示例：

• 在回顾《效果层级模型》中的三个序列后，该研究指出：“我们对序列并不特别关心，而是关心这些模型的前两个组成部分——知识和态度——对行为结果——投票——的影响力。换句话说，我们试图确定的并不是哪个先出现，而是哪个有更强的影响力：对问题的认知（第一层级）还是对属性的情感（第二层级）。”31

• 在回顾名人在Twitter上的使用和信息来源可信度之后，作者们表示：“认识到名人在社交媒体中产生的信息的新颖性和重要性，本研究测试了名人基于Twitter的口碑对消费者对信息来源可信度、传播电子口碑意愿、与品牌相关的结果、在线桥接社交资本以及与名人的社会认同感的影响。”32

• 在回顾归因风格之前，该研究指出：“本研究深入探讨了消费者不同的归因风格，具体研究了不同的归因风格如何影响对新消费品创新的评价。”33(2023)

• 在一项关于数学学习的研究中，作者回顾了相关文献并总结了研究结果，然后将其与该研究所设计的问题联系起来，如下所述：“简单来说，随着数学焦虑的增加，数学成就会下降。然而，焦虑和成就之间可能存在内在关系，一个明显的问题是：数学评估/问题上的差强人意表现是由于数学焦虑还是由于对内容的掌握不足？”34

• 在回顾对帮助学生学习英语有效的不同方法的研究后，该研究概括如下：“对于英语学习者来说，技术丰富的课程的显著特点和优势的检验强调了计算机辅助语言学习在第二语言教学中的关键作用。”35

• 在总结文献之前，一项测试退出过度拥挤市场的企业的研究的作者们注意到了这些研究之间的相似之处，并指出他们的研究填补了空白：“尽管在博弈论文献中，退出竞争有着重要的地位，但与之直接相关的实验研究却相对较少。”36

通过以上的示例，你可以看到这些研究者如何在他们的实验中明确地将其他研究与自己的研究联系起来，并将自己的研究放置在更大的研究背景中。这种明确的联系有助于读者理解为什么具体的研究被回顾，并展示了它们与当前研究的联系。

水平的一级相关性相似；没有一个达到了最低水平，甚至没有达到一级效果的平均水平，该平均水平在一级效果的元分析中为0.53。”30 这是一个将所有刚刚审查的研究与当前研究联系起来的句子的例子。

这些联系永远不会显得太过明显。良好的连接句包括：“这与我们的研究相关，因为……”、“我们在研究中利用这一发现，将其作为中介因素……”、“这些研究引发了关于……的问题，而这正是本研究的目的”，以及类似的表达。存在许多良好的连接句的示例；其中一些在"How To Do It"框3.3中提供。在阅读其他研究时，寻找这些连接句并适应到自己的工作中。

概念定义与操作化定义

文献综述也是研究中所有重要概念和变量在概念上进行定义的地方。概念定义是使用一般的定性术语对某个事物进行抽象的理论描述。学生们经常将概念定义与操作定义（或操作化定义）混淆，而操作定义则将概念以具体、具体的方式进行定义，以测量其具体特征。概念是关于某个事物的一般想法，具有许多具体特征。例如，在一项研究中，可信度在概念上被定义为“公众对新闻质量的感知”37，然后通过十三个具体特征对其进行操作化定义，其中包括新闻报道的公正性、完整性、准确性、可信度、可信度、信息性、有趣性、受欢迎程度、深度、重要性和写作质量，以及信息源和信息的可信度。这是两种不同类型的定义，它们出现在论文的两个不同部分。仅提供一系列用于测量概念的术语列表并不足以作为概念定义。概念定义应该基于理论描述和解释概念。另一个例子是道德判断的概念定义，即人们“用来判断一种行为在道德上是对还是错的理由”38。方法部分中给出的操作化定义是关于在面对特定困境时主题认为重要的事物的十二个陈述（有关示例，请参见脚注ii）。

For Review-No Commercial Use(2023)

ii这十二个陈述根据每个困境的不同而有所变化，以下是一个例子：(1) 这样的照片在一个职业生涯中只会出现几次；(2) 孩子们长大后会一直拥有这个可怕的、生动的提醒；(3) 这是一份家庭报纸，孩子们可能会在早餐桌上看到这张照片；(4) 你可能会接到很多愤怒的电话，人们会取消订阅；(5) 当这些人和他们的家人看到这张照片时，他们会有何感受；(6) 你的竞争对手正在进行类似的报道。如果你不刊登这张照片，你的竞争对手就会刊登类似的内容；(7) 发布这张照片将有助于你的报纸在调查报道方面的声誉；(8) 作为记者，我们是否有责任展示所有的新闻，无论情况如何；(9) 公众是否有权了解有关药物使用及其对人们，尤其是儿童的影响的所有事实；(10) 如何最好地服务社会；(11) 对某些人来说，痛苦的照片必须被展示，以使其他人受益；(12) 如果我不刊登这张照片，我可能会阻止这些孩子被夺走父母，但导致这种情况的条件将继续存在。

之前引用的可信度研究提供了一个很好的例子，说明不同研究者对一个概念的定义可能存在差异，其中包括将其定义为“对沟通来源的态度”，“对消息来源可信度的整体评估”，包括可信度、专业性和善意在内的组合，以及对媒体的质量、盈利能力、隐私、社区福祉和可信度的综合公众感知。对于某些概念的概念定义缺乏共识并不罕见。在这种情况下，重要的是在文献部分承认这一点，回顾不同的概念定义，然后捍卫你选择的定义并解释为什么它最适合你的研究。重申一下：在研究中使用的每个变量必须出现在三个位置：

• 在文献综述中进行概念定义，

• 在假设或研究问题中加以运用，以及

• 在方法部分进行操作化。

追踪每个变量贯穿整篇论文，确保每个变量都在这三个位置出现。不应该有孤立的概念或变量。

一个很好的方法是将文献综述视为一个大纲，列出所有的因变量、自变量、中介变量（中介者、调节变量、因果机制）和协变量（如果有的话）。使用与大纲相对应的副标题来帮助读者理清思路，也有助于组织写作。这样确保了所有内容都有定义，涵盖了经验证据，并且关于一个变量的所有讨论都在同一个地方。经验丰富的读者期望在同一个概念中找到所有相关信息，并且当某个概念在论文中的不同位置出现时可能感到不满。

涉及多个实验的文献综述

许多研究人员在一篇文章中报告了多个实验，原因包括复制发现和排除合理的替代解释。有时这是事先计划的，而其他时候则是在第一项研究结果知道后才出现的需求。这个现象将在第5章关于有效性中探讨。当一篇论文中报告多个实验时，可能有一个部分涵盖了两个研究的所有文献，或者可能有两个文献综述，第一个最长，第二个只包括与第二个研究相关的新文献。通常，它们将被标记为“研究1”和“研究2”，如果有更多的话还会有更多的标记。

实验。一般的模式是先报告第一个研究的文献综述，然后在类似于“研究1”的子标题下报告其方法，然后是结果和讨论。第二个研究将在“研究2”或“实验2”的子标题下进行，首先描述该研究，如果需要阐明任何新变量，可以进行简短的文献综述。例如，如果为了排除第一个研究中可能出现的混杂因素或合理的替代解释，引入了新的自变量、中介变量或调节变量，那么它们显然不会在第一个文献综述中涵盖。不需要重复主要文献综述中的所有先前信息，只需添加新的概念并将其与先前的文献和理论联系起来。一个很好的例子来自一项关于议程设置中认知加工的研究。第二个实验是在第一个实验显示出作者预期的效果在某些受试者上有作用而在其他受试者上没有作用之后构思的。他们推测这可能受到了他们使用的故事内容的影响，并写道：“我们决定进行后续研究，以测试这样一个假设，即相较于其他人更加中心地加工信息的人可能会受到文章内容方面的影响，例如记者对问题重要性的评估。”然后他们用“研究2”作为子标题，并以一页纸的文献综述开始，该文献综述仅涉及处理新闻故事内容的研究。

最后，还有一些来自文献综述的其他重要提示：

• 包括最新的文献来源。除了经典或重要的研究，确保包括最新的作品。如果研究是在一段时间之前开始的，这一点尤为重要。回顾一下最近发表的期刊，看看是否有遗漏的研究。在这个过程中，包括计划提交的期刊。

• 使用不同的搜索方法。使用数据库中的关键词进行搜索并不能总是找到所有重要的文献，即使使用多个数据库。为了避免遗漏重要的研究，浏览过去两三年可能会发表关于该主题的文章的期刊，看看是否有遗漏的内容。在这个过程中，包括计划提交的期刊。

• 包括相互矛盾的发现。很少有研究项目始终产生一致的结果。寻找那些与预期不符的研究。不仅诚实是重要的，了解可能导致你的实验不按计划进行的原因也很重要，这样你就可以尽可能地进行控制。回顾那些不支持你的预测的研究，并解释为什么它们可能以这种方式出现。

• 适度引用他人的观点。过多或过频繁地引用他人的观点会引起读者的警惕，因为这可能会让人觉得你没有理解所引用内容的

不要滥用引用。除非没有更好、更清晰、更简洁的表达方式，否则要养成不使用直接引语的习惯。学术论文中符合这一标准的句子很少。使用直接引语时，请适当引用，并注明页码。

• 注意避免抄袭。请参考前面的观点。许多抄袭行为是由于粗心而不是恶意导致的，例如，通过从原始文章中复制和粘贴来做笔记，然后忘记这些笔记是否已经改写成自己的话。从其他文章中重复自己的写作也是一种抄袭行为。建立一个系统，了解哪些是改写的内容，哪些不是，例如通过高亮显示、使用引号或使用不同的字体。

• 始终阅读论文。追溯并阅读其他人文献综述中引用的原始论文可能需要耗费时间，但依赖于他人对论文的解释而不是自己阅读可能是危险的。学者们有时会不加批判地一遍又一遍地报告同一项研究的摘要，甚至可能是错误的。42我发现这种情况尤其常见，特别是在书籍或其他文章的文献综述中。

#### 假设与研究问题

文献综述部分也是假设和研究问题通常被纳入的地方。一种方法是在文献综述的末尾列出所有的假设。这种风格涉及在方法部分之前使用一个假设模块，如下所示：

“假设1（H1）：当个人捐赠者知道非营利组织接受一定数量的政府资金时，他们的捐赠金额会减少。（分类排挤假设）

假设2（H2）：当非营利组织从政府获得更大比例的资金时，个人捐赠者的捐赠金额会减少。（连续排挤假设）

假设3（H3）：知道政府资金来自竞争性的基于功绩的计划将增加个人捐赠的意愿。（排挤内化假设）”43

术语“捐赠”和“给予”可能具有不同的定义，因此在编写假设时重要的是不要使用同义词。尽量使用平行的措辞来表达假设。

为了避免读者在回顾整个文献综述以查找每个假设的证据时不得不翻阅许多页面，一些期刊在假设之前或之后提供了一份关于每个预测所依据的文献的简要摘要。在文献综述部分详细描述研究时，再在结尾重复一份摘要可能效率不高，另一种方式是将假设嵌入到与各个假设相关的文献综述部分中。例如，关于气候观念判断的研究比较了明示线索和隐含线索的相对权重，44作者首先回顾了明示线索并提出了以下假设：

“H1：调查作为明示媒体线索会影响接收者对气候观念的判断，使其朝着调查结果的方向发展。”45

接下来，他们回顾了隐含线索，并提出了以下预测：

“H2：如果媒体报道中的论点支持（反对）某种观点，则公众对该观点的认同会增加（减少）。46”

然后，他们回顾了媒体中这两种线索的说服效果的文献，并给出了以下假设：

“H3a：如果提供有关未知问题的调查信息，接收者倾向于追随多数意见。”47

还有几个假设，所有这些都在模型中以图表形式总结（见图3.1）。

在假设之前引入概念而未在文献综述中详细阐述该概念可能是一项挑战。例如，不要在文献综述中未讨论过某个概念的情况下引入包含该概念的假设。



**FIGURE 3.1**

**A MODEL SHOWING THE IVS, DVS, AND PROCESS**

**PROPOSED IN HYPOTHESES**

H3a

H1

H3b

**Perceived climate of opinion**

**Slant of argumentation**

(Implicit cue)

**Tendency of poll results**

(Explicit cue)

|  |  |
| --- | --- |
| **Personal opinion** | |
|  | H4 |

H2

来源：Zerback，Thomas，Thomas Koch和Benjamin Kramer。2015年。“思考他人：隐性和显性媒体线索对舆论氛围感知的影响。”《新闻与传播季刊》92（2）：421–443。

在探讨概念之前就引入假设可能会让组织文献回顾变得棘手，因为它并不假设读者对这些概念有先前的了解。这使得将所有的假设都放在最后成为一种诱惑，但是读者在到达H1时可能已经很难回想起关于这些概念的内容。相反，将假设分解成较小的陈述，一个自变量一个，按照过程发生的顺序进行，这是关键。

在前面的示例中，假设被正式表述，并带有“H”加上一个数字的标识（H1，H2，H3等）。小写字母（H5a，H5b，H5c）表示相关假设的子集。但是，假设以叙述风格表达也非常常见。例如，在一项关于威权政府是否愿意将公民偏好纳入他们制定的政策中的研究中，中国的学者以以下方式提出他们的假设：

因此，根据以前的研究以及中国共产党的声明，我们有理由相信当国家与社会关系和谐时，领导人可能对通过互联网和正式渠道表达的意见同样持开放态度。然而，当领导人认为国家与公民之间存在对立时，我们预计他们将对正式渠道和互联网渠道的意见持较少开放态度。48

假设应该用将来时态（例如，“将会有”，“可能是”），而不是现在时态（例如，“是”），因为它们尚未经过测试；现在时态听起来像是已经成为事实。current or future state of something until the empirical evidence on that has been covered.

需要多少个假设？学生经常问这个问题，但我只能给出“金发女孩”式的回答：不要太多，也不要太少，恰到好处。一个假设可能表明研究规模太小，或者需要将一个过于庞大的假设拆分为更小的部分。超过十个假设可能意味着研究试图做得太多，应该分成两个独立的出版物。让理论、逻辑和读者的注意力来决定。让同事们阅读文献回顾，看看他们是否能够清楚地理解，而不会迷失或感到厌倦。

比假设的数量更重要的是它们的清晰性。如果指定了一个过程，假设应该在逻辑上相互补充。有时，实验探索的因果机制涉及复杂的理论过程中的多个步骤。在这种情况下，将事物分解为一系列特定的假设，并建立一个连贯的模型是有意义的。例如，在研究照片如何提升道德判断的研究中，该过程被分解为三个假设：

“H1：看到照片的参与者在利益相关者方面的详细程度明显高于没有看到照片的参与者。

H2：道德推理与利益相关者的详细程度显著相关。

H3：看到照片的参与者的道德推理水平明显高于没有看到照片的参与者。”49

第三个假设是研究最感兴趣的，但它在因果事件链中处于最后。

Mutz在关于近距离镜头视角和不文明电视言辞的研究中以类似的叙述方式提出了一组假设（在此处为了阅读方便，它们被分成了单独的段落）：

“第一个假设是近距离镜头视角和不文明行为都会增加情绪激发的程度。

第二个假设是，由第一个假设得出，情绪激发的增加会增加回忆的程度。

iii该研究还预测了照片和参与度之间的互动效应，但我将把主效应和互动效应的讨论留到后面。

在这个例子中，第一个预测涉及近距离摄像透视和不礼貌行为导致情绪激发。第二个预测将来自第一个假设的激发与回忆联系起来。顺序并不是随机选择的；研究者按照这个顺序提出它们，是因为假设它们按照这个顺序发生。对于更复杂的过程，可以通过图表来展示这些路径，以便更容易进行可视化。

假设基础

零假设与备择假设

在解释如何撰写假设之前，让我们回顾一下零假设和备择假设的概念。零假设表明没有差异或方向性，备择假设表明存在差异或方向性。这些概念的提出是为了减少实验者的偏见，或者说是为了避免寻找支持自己观点的证据。很久以前，科学家们提出了一种基本上反对自己立场的观点，也就是研究者不应该试图证明某种处理引起效应，而是试图证明它没有。换句话说，试图证明两个事物之间没有关系或两个组之间没有差异。零假设用H0表示，备择假设用Ha或H1表示。通过否定零假设，我们可以认为存在关系或差异的可能性，因此我们可以说我们的假设是正确的。技术上讲，这就是备择假设。换句话说，我们从零假设（无差异或无关系）为真的立场出发，直到有证据表明相反为止。如果拒绝零假设，我们说我们找到了支持我们的假设的证据。我们从不使用“证明”这个术语，因为技术上是不可能的。相反，应该使用适当的语言来支持我们的假设或拒绝零假设；拒绝假设或“未拒绝”；或者说假设“未被否定”。这在研究论文中很少明确说明，但却是所有论文背后的思维过程。

假设是独立变量和因变量之间关系的预测。可以将其视为“如果...那么...”的陈述，例如，“如果记者看到照片，那么他们将使用更高水平的道德判断”。假设应该以概念为基础，而不是具体的操作定义。例如，在商业中，有一个概念称为管理信任度（MTW），在概念上定义为“对主管的信任度”。通过九个项目进行测量，包括能力、胜任、关心和强烈的正义感等。假设涉及管理信任度，而不是胜任、正义等。例如：

“H1: 强调与设定具有挑战性但可行的目标相关的内部管理对个体对MTW的感知有积极影响。”

**差异和方向的区别**

假设分为两种类型：一种是预测因果方向的假设，即哪个变量的影响比另一个更大；另一种是仅预测差异而不提出哪个变量将比另一个更大。后者也被称为无方向性假设。一个预测差异的假设的例子是：

H: 看到照片的记者在道德判断上会与未看到照片的记者显著不同。

这个假设预测了看与不看照片之间的差异，但没有指明结果会更高或更低。根据这个假设，看到照片可能会导致道德判断更好或更差。相比之下，这个假设的另一个版本预测了一个方向：

H: 看到照片的记者在道德判断上会显著比未看到照片的记者更高。

第二个假设明确指出了两种条件之间的差异，即看与不看照片的差异。上述管理信任度研究的另一个方向性假设的例子是：

“H1: 与设定具有挑战性但可行目标相关的内部管理强调对个人对管理信任度的感知将产生积极影响。”56

该假设使用了“积极”一词，预测了一个因果方向，即研究者预期目标设定条件下的管理信任度水平更高。一种无方向性的表述方式是，目标设定会对管理信任度产生不同于其他条件的影响，但不指明该影响是积极还是消极。

选择预测方向或差异的假设取决于可用的理论和经验证据。如果有足够的证据表明因果方向，那么预测这个方向是合适的。如果没有足够的证据，或者理论不提供方向性预测，则更适合预测差异而不指定方向。例如，在第一次研究照片对道德判断的影响时，根据理论和经验证据的限制，可能更适合预测差异而不预测具体的方向。

在测试道德判断方面的照片效应时，存在理论上的原因使人们相信照片可能会改善或恶化道德判断。在这种情况下，合适的做法是提出看到照片会导致被试者行为不同的假设，而不预测他们的道德判断是否比那些未看到照片的人更好或更差。在研究发现某个方向的效应之后，从那时起，使用方向性假设是合适的。

如果理论和先前的证据非常有限，以至于研究者无法进行任何预测，那么应该使用一个研究问题。研究问题代表了关于理论和证据的更初步的情况。因此，继续上面的例子，如果没有足够的理论和证据来预测照片对道德判断会产生某种影响，那么提出问题“看到照片是否会影响记者的道德判断？”是合适的。作为一个经验法则，根据现有的证据和理论进行尽可能具体的预测是最好的，因为这为我们提供了更强大的分析工具。

研究者是否对方向性进行预测对于如何在统计上测试假设是重要的。对于那些没有接受过统计学入门课程的读者来说，下面的内容可能没有意义；对他们来说，重点在于所使用的假设类型在统计测试中很重要。对于基本统计学知识的读者来说，方向性假设适合使用单尾检验。对于无方向性假设，使用双尾检验。（为简单起见，这些例子假设只有两组，即一个处理组和一个对照组。）这在正态分布图中可以找到图形化的说明。无方向性或差异性假设使用双尾t检验进行测试。因为没有指定方向，只要t值落在正态分布的任一端，即图3.2中蓝色阴影区域内，就支持这个假设。

只有当t值落在假设指定的正态分布的一端时，才支持方向性假设。在这个例子中，假设是照片会导致道德判断水平更高，所以t值必须落在图3.3中表示的蓝色阴影区域内才具有显著性。实际上，如果存在显著性，使用方向性假设更容易找到它；然而，如果预测的方向是错误的，那么可能得到错误的结论。



**FIGURE 3.2 NONDIRECTIONAL HYPOTHESIS**

**±1.96 Standard Deviations**

0.025

0.95

0.025

3 2 1 0 1 2 3



**FIGURE 3.3 DIRECTIONAL HYPOTHESIS**

##### For Review-No Commercial Use(2023)

This right area shaded dark gray is .05

of the total area under the curve.

0 1.645

**Normal Probability**

如果统计测试发现显著性，但结果与预测的方向相反，那么假设就不被支持。

假设写作公式

既然这本教科书已经讲解了何时以及为什么要提出明确规定方向的假设，我提供一个基本的公式来编写明确而重点明确的假设。对于每个假设，该公式包括以下内容：

• 预测要么是具有统计显著差异的预测，要么是差异的方向（用下划线表示）。

• 自变量（用粗体表示）。

• 因变量，以概念术语表述（用双下划线表示）。

对于使用两组实验，即处理组和对照组的实验，假设使用“between”一词。使用将来时态也更为常见。可以类似填空的方式书写如下：

• 看到Twitter上的报道和在网站上看到的报道之间的可信度评分（因变量）将有显著差异（自变量）。58

• 看到长篇或短篇报道（自变量）的受试者对重要性评分（因变量）将有显著差异。59

• 暴露于倡导框架（自变量）的观众会将犯罪的重要性（因变量）与暴露于客观报道（自变量）的观众显著不同地排名。60

这些假设都没有明确指出差异是更高还是更低，更大还是更小。它们也可以以指定方向的方式书写，例如：

• 看到Twitter上的报道的可信度评分（因变量）将显著高于在网站上看到的报道的评分（自变量）。

• 看到短篇报道的重要性评分（因变量）将显著低于看到长篇报道的评分（自变量）。

• 暴露于倡导框架（自变量）的观众将将犯罪的重要性（因变量）显著提高于暴露于客观框架（自变量）的观众。

如果一个实验涉及三个或更多组，例如两个处理组和一个对照组，使用“among”而不是“between”，如下所示：

• 看到Twitter上的报道、网站上的报道或报纸上的报道的可信度评分（因变量）之间将有显著差异（自变量）。

请注意，这个例子只是将自变量改为包括三个条件——Twitter、网站和报纸，而不是两个。

在所有这些基本假设的例子中，每个假设都被准确地表述。它们使用清晰、简单、单一的因果预测。使用的是相同的术语而不是同义词。这些术语应该与文献综述中使用的术语相同。每个假设都预测了特定自变量和因变量之间的明确因果关系。如果可以通过理论和/或先前的证据支持，更喜欢使用有方向性的假设。

如果实验中包括协变量或统计控制变量，这些变量已被证明影响因变量，但在随机分配中分布不均等，则在假设末尾添加“控制年龄、教育和性别”等变量的短语：

H1：看到Twitter上的报道、网站上的报道或报纸上的报道之间的可信度评分将有显著差异，在控制年龄、教育和性别的情况下。

多个自变量的假设

到目前为止，在本书中，我们已经讨论了逐个操纵一个自变量的情况，例如在前面的假设中，信息的格式是被实验对象所看到的Twitter、网站或报纸。在这个例子中，独立变量是框架：宣传性或客观性。这些被称为单因素设计，并将在后面的章节中详细解释。然而，在实际研究中，许多实验会同时操纵多个独立变量，例如在同一研究中操纵信息的格式和框架。当操纵超过一个独立变量或因素时，应该为每个因素编写单独的假设，也称为主效应，以及两个因素的交互作用或考虑两个因素一起产生的效应。因此，例如，在一个操纵信息格式（Twitter、网站）和框架（宣传性、客观性）的实验中，应该有两个假设，分别针对每个独立变量或因素的主效应：

H1：读Twitter上的报道的受试者的可信度评分将显著高于读网站上的报道的受试者的可信度评分。

H2：读客观框架的报道的受试者的可信度评分将显著高于读宣传性框架的报道的受试者的可信度评分。

最后，还应该有一个关于信息格式和框架之间交互作用的假设或研究问题，例如：

H3：读网站上的客观框架报道的受试者的可信度评分将高于其他信息格式和框架组合的任何情况；

或者

RQ1：信息格式和框架之间是否存在交互作用？

当然，选择是预测还是提出研究问题取决于理论和证据；尽可能使用假设。下一个交互作用假设的例子展示了两个实验操作或独立变量——恶心感和伤害感——与受试者先前的政治观点相互作用。作者预测，根据受试者是否支持或反对某一政治做法，独立变量的效应将不同：

“H4：在反对一项政治做法的人中，偶发的恶心感和伤害感联系将增加道德坚定和导致更严厉的道德判断，而在支持者中将产生相反的效果。”62

对于涉及多个独立变量或因素的研究，这些假设还包含了一个逻辑的推进，从一个自变量到另一个自变量，然后到两个自变量的相互作用。如果提出了中介变量，那么它们也会按照逻辑顺序逐个考虑。为了更清晰地思考这些更复杂的假设，创建一个表格列出独立变量或因素及其水平、中介变量和因变量可能会有所帮助。可以使用这些表格来可视化假设，如图3.4所示。

下面是一个包含中介变量的假设的例子：权力感概念。这个实验研究了应对食品污染组织危机的发言人。这个实验中的独立变量是声音音调，有高音调或低音调两个水平。中介变量是权力感，它在低音调的声音中被感知到。因此，假设是：

“声音音调降低会导致对能力的更高认知，与声音音调升高相比，这是由于被感知到的权力感的中介效应（H2）。”62

在发表的期刊文章中，假设的变化还有很多；这旨在向初学者介绍假设写作的基础知识。在阅读已发表的实验和开展自己的研究时，请根据每个研究的需要调整你的假设写作方式。

倡导目标

**如何操作 3.4**

假设编写表格

编写假设时，为了更容易地绘制出步骤，可以通过以下表格来可视化变量。在左侧的列中列出独立变量或因子，并列出每个变量的水平。在右侧的列中列出因变量。如果存在中介变量或因果机制，请在中间列中列出它们。然后，绘制线条来展示将成为假设的过程和预测。下面的表格是简化的，但可以根据需要扩展以适应任意多的因子、中介变量和因变量。

**当存在两个或多个因素时的交互作用效应**

**因变量**

**介入变量或因果机制**

**独立变量或因素及其水平**

H1

平台

*Twitter 网站*

H3 可信度

框架

H2

H1: 在 Twitter 上阅读故事的受试者的可信度得分明显高于在网站上阅读故事的受试者。（平台因素的主要效应。）

H2: 阅读客观框架故事的受试者的可信度得分明显高于阅读倡导性框架故事的受试者。（框架因素的主要效应。）

H3: 阅读网站上的客观框架故事的受试者的可信度得分将高于其他消息格式和框架的组合。（平台和框架因素的交互作用。）

**对于基于彼此构建的因果机制的假设**

H2

**因变量**

**介入变量或因果机制**

**自变量或因素及水平**

照片

看到

未看到

*(Continued)*

伦理推理

H3

对利益相关方的详细阐述

H1

(继续)

H1: 观看照片的参与者在利益相关方方面的详细阐述程度将明显高于未观看照片的参与者。（因素导致因果机制。）

H2: 对利益相关方的更多详细阐述将与更高的伦理推理水平显著相关。（因果机制导致因变量。）

H3: 对于观看照片的参与者，其伦理推理水平将明显高于未观看照片的参与者。（因素导致因变量。）

在撰写假设时，最后一条建议是要思考每个假设的重要性。问问自己这些假设是否有趣、有意义或重要。它们是否不言而喻或已经经过充分测试？尽量避免提出我所谓的“显而易见的假设”，即那些答案显而易见的假设。例如，媒体最关注的议题与公众认为最重要的议题是相同的这个基本议程设定假设已经经过充分测试。没有必要把这个问题当作第一次提问一样去测试，不会有什么惊人的发现。许多议程设定的研究从未提出这个问题。当它们确实提出时，通常是为了更具体的预测，或者看看未经测试的条件是否会导致与过去四十多年反复显示的结果有所不同。

还要问问如果以下三种情况之一发生，会意味着什么：

• 如果假设得到支持，那意味着什么？

• 如果没有支持（没有显著性），那意味着什么？

• 对于定向假设，如果显著性发现与预测的方向相反，那意味着什么？

这个练习旨在帮助你思考所有可能的结果以及每种结果的含义。这将帮助你思考研究的意义，以及你如何在讨论部分解释结果。如果在这些情况下你无法想象出结论，或者它们似乎显而易见或不重要，那么就是重新考虑研究的时候了。此外，考虑在研究中可能包括哪些证据来支持解释。例如，如果数据不能支持假设，或者发现显著性与预测相反，那可能是什么原因？这个练习还涉及到我经常在初学者中发现的一个现象。

实验研究者往往会感到自己的假设得不到支持就“失败”了。即使在不能支持假设的研究中，仍然能够学到一些有价值的东西。在某些情况下，不支持一个假设比支持它更有趣。毕竟，进行研究的目的首先并不仅仅是支持所有假设，而是为了获得有价值的发现。正如坎贝尔和斯坦利所说，“理论测试的任务主要是拒绝不充分的假设。”请放心，不能支持假设并不意味着研究不能被发表。

研究问题

除了用于探究两个自变量之间的相互作用，研究问题有时也用于主效应。当没有足够的证据或理论支持一个预测时，或者在测试模型时，会提出这些问题。在实验中编写研究问题与其他方法并无不同。除了之前展示的正式研究问题风格外，研究问题也可以用叙述风格来书写，例如：

总的来说，本研究涉及四个主要问题。第一，这两种教学干预在CVS学习的整体效果上是否有差异？第二，这两种教学干预在学习转化到新领域的程度上是否有差异？第三，学生从这两种干预中学到的内容是否有所不同？具体来说，他们是否从基于无效设计的干预中更好地理解控制无关变量的必要性？第四，这两种教学干预在不同学习环境中的学生（例如成绩较高与较低的学校）之间是否有不同的相对效果？64

设计实验的过程不一定按照本书的章节顺序进行。有时，研究人员需要重新审视已经完成的事情。编写良好、清晰的假设就是其中之一。此外，通常需要写、改写、编辑、再写、再编辑、再重写，直到对假设的准确性和清晰性感到满意。

下一章将讨论实验设计，包括本书重点关注的真实实验的经典版本，以及准实验、自然实验和现场实验。与假设编写一样，一旦您思考了自己的设计，可能需要回过头来重新修改假设和研究问题。这是正常的，研究是一个杂乱的过程。

**82** Designing Experiments for the Social Sciences

* Not knowing the relevant literature thoroughly enough
* Orphan Variables—not having all the variables conceptually defined in the hypotheses or research questions, and operationally defined in the methods sections
* Not explaining the posited relations among variables and how they relate to theory

**常见错误**

###### Test Your Knowledge

1. 下列哪个是概念定义的例子？

a. 威权主义被定义为一种与不确定性厌恶和对威胁敏感性强相关的个性特征。

b. 威权主义通过尊重长者、有良好礼貌、服从和表现良好来进行衡量。

c. 威权主义通过一个7分制度量。

d. 威权主义在不同研究者间有不同定义。

2. 下面的假设属于哪种类型？

“具有高水平专业知识的来源将对可信度的感知产生积极影响。”

a. 定向假设

b. 无方向假设

c. 交互作用假设

d. 零假设

3. 是否制定定向或无方向假设取决于你愿意冒多大的风险，即你的假设是否会得到支持。

a. 真

b. 假

4. 当你有多个假设时，总是从代表你正在测试的过程的最终结果开始。

a. 真

b. 假

Chapter 3 ■ Theory, Literature, and Hypotheses **83**

5. 文献综述的目的是 \_\_\_\_\_\_。

a. 发展理论贡献

b. 解释变量的测量方法

c. 提供足够的细节，使其他研究人员能够复制研究

d. 在前人的研究基础上进行进一步构建

6. 实验的想法应该来源于：

a. 直觉和传统智慧

b. 反复试验和错误的过程

c. 理论和现有证据

d. 以上都不是

7. 假设应该 \_\_\_\_\_\_。

a. 被描述为已被证明或未被证明

b. 用具体操作化的术语陈述

c. 放在方法部分

d. 在逻辑上相互关联

8. 当没有足够证据进行预测时，应该使用研究问题代替假设。

a. 真

b. 假

9. “看到照片的记者在道德判断上将与未看到照片的记者显著不同。” 这个陈述属于以下哪个选项？

a. 定向假设

b. 无方向假设

c. 研究问题

d. 零假设

10. 是否制定定向或无方向假设取决于 \_\_\_\_\_\_。

a. 理论

b. 证据

c. 使用的统计检验方法

d. 既有理论又有证据

**84** Designing Experiments for the Social Sciences

答案:

1. a

2. a

3. b

4. b

5. d

6. c

7. d

8. a

9. b

10. d

1. 从第一章的任务中选择你的三个想法之一，并在其中增加至少五页的文献综述，包括发展、测试或拓展该理论以及相关文献。至少使用25篇文章。确保对文献进行分析和综合，而不仅仅是总结，并将文献与你的研究联系起来。
2. 为你正在开发的实验编写两到三个明确而简明的假设。使用本章的公式。如果你的实验中有多个因素（自变量），为每个因素的主效应编写一个假设，并为交互作用效应编写一个假设。让两位同事阅读并检查是否清晰明了。

**应用练习**

Chapter 3 ■ Theory, Literature, and Hypotheses **85**

1. R. Barker Bausell, 《Conducting Meaningful Experiments: 40 Steps to Becoming a Scientist》（千橡市，加利福尼亚州：Sage，1994年），32页。

2. Thomas S. Kuhn, 《The Structure of Scientific Revolutions》，第3版（芝加哥：芝加哥大学出版社，1996年）；Earl Babbie，《The Practice of 《Social Research》第11版（加利福尼亚州贝尔蒙特：汤普森-沃兹沃斯出版社，2007年）

3. H. W. Turnbull，《艾萨克·牛顿通信：1661–1675年，第1卷》（伦敦：The Royal Society at the University Press，1959年）。

4. Esther Thorson, Robert H. Wicks 和 Glenn Leshner，《新闻与大众传播研究中的实验方法论》，《新闻与大众传播季刊》89卷，第1期（2012年）：112–124页。

5. Rebecca B. Morton 和 Kenneth C. Williams，《实验政治科学与因果关系研究：从自然到实验室》（纽约：剑桥大学出版社，2010年）。

6. Paul D. Leedy 和 Jeanne Ellis Ormrod，《实用研究：规划与设计》（波士顿：Pearson Education Inc.，2010年），第5页。

7. Kurt Lewin，《社会科学中的场论：库尔特·莱文选定的理论论文》（伦敦：Tavistock，1952年），169页。

8. Leedy 和 Ormrod，《实用研究》。

9. S. Iyengar，《有责任吗？：电视如何塑造政治议题》（芝加哥：芝加哥大学出版社，1991年）。

10. L. R. Franklin，《探索性实验》，《科学哲学》72卷，第5期（2005年）：888–899页；C. Kenneth Waters，《探索性实验的性质和背景：引介三个探索性研究案例研究》，《生命科学的历史与哲学》29卷，第3期（2007年）：275–284页。

11. Franklin，《探索性实验》；Waters，《探索性实验的性质和背景》。

12. Franklin，《探索性实验》。

13. 同上。

14. Babbie，《社会研究实践》。

15. Bausell，《有意义的实验进行》。

16. 同上。

17. Renita Coleman，《视觉对道德推理的影响：照片对新闻从业人员在道德决策中的价值是多少？》，《新闻与大众传播季刊》83卷，第4期（2006年）：835–850页。

18. H. B. Brosius，《电视新闻中情感图片的影响》，《传播研究》20卷，第1期（1993年）：105–124页；David Domke, David Perlmutter 和 Meg Spratt，《时代的代表？：视觉图像的“力量”检验》，《新闻学》3卷，第2期（2002年）：131–159页；A. Friedman，《图片构架：知识在自动编码和记忆要义中的作用》，《实验心理学通论》108卷（1979年）：316–335页；Doris Graber，《在记忆中的视觉：视觉图像如何促进从电视新闻中学习》，《传播学杂志》40卷，第3期（1990年夏季）：134–155页；George E. Marcus, W. Russell Neuman 和 Michael MacKuen，《情感智能与政治判断》（芝加哥：芝加哥大学出版社，2000年）；Richard E. Petty 和 John T. Cacioppo，《沟通与说服：态度变化的中心和周边路径》（纽约：斯普林格出版公司，1986年）。

19. Elinor Amit和Joshua D. Greene，《你看，目的并不能辩解手段：视觉形象与道德判断》，《心理科学》23卷8期（2012年）：861-868；J. D. Greene等，《道德判断中的认知冲突和控制的神经基础》，《神经元》44期（2004年）：389-400；J. D. Greene等，《道德判断中情感参与的fMRI研究》，《科学》293期（2001年）：2105-2108；Joshua Greene和Jonathan Haidt，《道德判断如何工作（以及在哪里）？》，《认知科学趋势》6卷12期（2002年）：517-523；D. Kahneman和S. Fredrick，《再次思考典型性：直觉判断中的属性替代》，收录于《启发式与偏见》一书中，编辑T. Gilovich、D. Griffin和D. Kahneman，（纽约：剑桥大学出版社，2002年），49-81页；Emma Rodero，《在广播故事中观看：声效和镜头对音频虚构作品中唤起的想象和注意力的影响》，《传播研究》39卷4期（2012年）：458-479；Adam B. Moore、Brian A. Clark和Michael J. Kane，《谁不可杀？工作记忆能力、执行控制和道德判断的个体差异》，《心理科学》19卷6期（2008年）：549-557。

20. Vicki Goldberg，《摄影的力量：照片如何改变我们的生活》（纽约：Abbeville，1991年）。

21. Bausell，《进行有意义的实验》，第21页。

22. H. Denis Wu和Renita Coleman，《推进议程设置理论：两级议程设置效应的比较优势和新的依赖条件》，《新闻学与大众传播季刊》86卷4期（2009年冬季）：775-789。

23. James Potter，《媒体暴力论》（千橡市，CA：Sage，1999年）。

24. Thorson、Wicks和Leshner，《实验方法论》。

25. Pamela Shoemaker、James W. Tankard和Dominick Lasorsa，《如何构建社会科学理论》（千橡市，CA：Sage，2003年），第11页。

26. 同上，第11页。

**86** Designing Experiments for the Social Sciences

27. George Gerbner等，《与电视共度时光：培养过程的动态》，收录于《媒体效应视角》一书中，编辑Jennings Bryant和C. Zillmann（Hillsdale，NJ：Erlbaum，1986年），17-40页；George Gerbner等，《与电视一同成长：培养视角》，收录于《媒体效应：理论与研究进展》一书中，编辑J. Bryant和C. Zillmann（Hillsdale，NJ：Erlbaum，1994年），7-14页。

28. 同上。

29. Coleman，《视觉对道德推理的影响》。

30. Wu和Coleman，《推进议程设置理论》，第778页。

31. 同上，第777页。

32. Seung-A Annie Jin和Joe Phua，《关注名人关于品牌的推文：基于Twitter的电子口碑对消费者信源可信度感知、购买意愿和与名人的社会认同的影响》，《广告学杂志》43卷2期（2014年）：183页。

33. Nan-Hyun Um和Wei-Na Lee，《文化是否影响消费者对负面名人信息的处理？文化在负面名人信息评估中的影响》，《亚洲传媒学》25卷3期（2015年）：329页。

34. Elena Novak和Janet Tassell，《利用电子游戏提高教育专业学生的数学表现：一项实验研究》，《人机行为计算机》53期（2015年）：125页。

35. Horacio Alvarez-Marinelli等，《在哥斯达黎加小学中进行计算机辅助英语语言学习：一项实验研究》，《计算机辅助语言学习》29卷1期（2016年）：105页。

36. Ryan Oprea、Bart J. Wilson和Arthur Zillante，《消耗战：基于市场退出的实验研究》，《经济调查》51卷4期（2013年）：2019页。

37. Miglena Mantcheva Sternadori和Esther Thorson，《匿名消息来源损害所有报道的可信度》，《报纸研究杂志》30卷4期（2009年）：56页。

38. James R. Rest、Lynne Edwards和Stephen J. Thoma，《设计和验证道德判断的测量工具：阶段偏好和阶段一致性方法》，《教育心理学杂志》89卷1期（1997年3月）：5-28页。

39. Sternadori和Thorson，《匿名消息来源损害所有报道的可信度》，56页。

40. Kristin Bulkow、Juliane Urban和Wolfgang Schweiger，《议程设置的二重性：信息加工的作用》，《公共舆论研究国际杂志》25卷1期（2013年春季）：43-63页。

41. 同上，52页。

42. Tara Halle，《一个警示故事：你检查过引文了吗？》，《健康报道》（2017年）；Lisa Marriott，《在实验研究中使用学生受试者：使用学生作为纳税人代理的做法的挑战》，《社会研究方法国际杂志》17卷5期（2014年）：503-525页。

43. Mirae Kim和Gregg G. Van Ryzin，《政府资助对艺术组织捐赠的影响：一项调查实验》，《非营利组织和志愿部门季刊》43卷5期（2014年）：913页。

44. Thomas Zerback、Thomas Koch和Benjamin Kramer，《关于他人的思考：隐性和显性媒体线索对舆论氛围感知的影响》，《新闻学与大众传播季刊》92卷2期（2015年）：421-443页。

45. 同上，423页。

46. 同上，424页。

47. 同上，425页。

48. Tianguang Meng、Jennifer Pan和Ping Yang，《对公民参与的有条件接纳度》，《比较政治研究》（2014年）：8页。

49. Coleman，《视觉对道德推理的影响》，839页。

50. Diana C. Mutz，《“当面”电视辩论对合法反对派认知的影响》，《美国政治科学评论》101卷4期（2007年）：624页。

,

Chapter 3 ■ Theory, Literature, and Hypotheses **87**

51. D. T. Campbell和J. C. Stanley，《研究的实验和准实验设计》（芝加哥：Rand McNally，1963年），35页。

52. Thorson, Wicks, and Leshner，《实验方法论》。

53. 同上。

54. Y. J. Cho和E. J. Ringquist，《管理信誉和组织结果》，《公共行政研究与理论杂志》21卷（2011年）：53-54页。

55. Mogens Jin Pedersen和Justin M. Stritch，《内部管理和感知管理信誉度》，《美国公共行政评论》（2016年）：5页。

56. 同上。

57. Thorson, Wicks, and Leshner，《实验方法论》。

58. 改编自Mike Schmierbach和Anne Oeldorf-Hirsch，《小鸟告诉我，所以我不相信：Twitter、可信度和问题认知》，《传播季刊》60卷3期（2012年7-8月）：317-337页。

59. 同上。

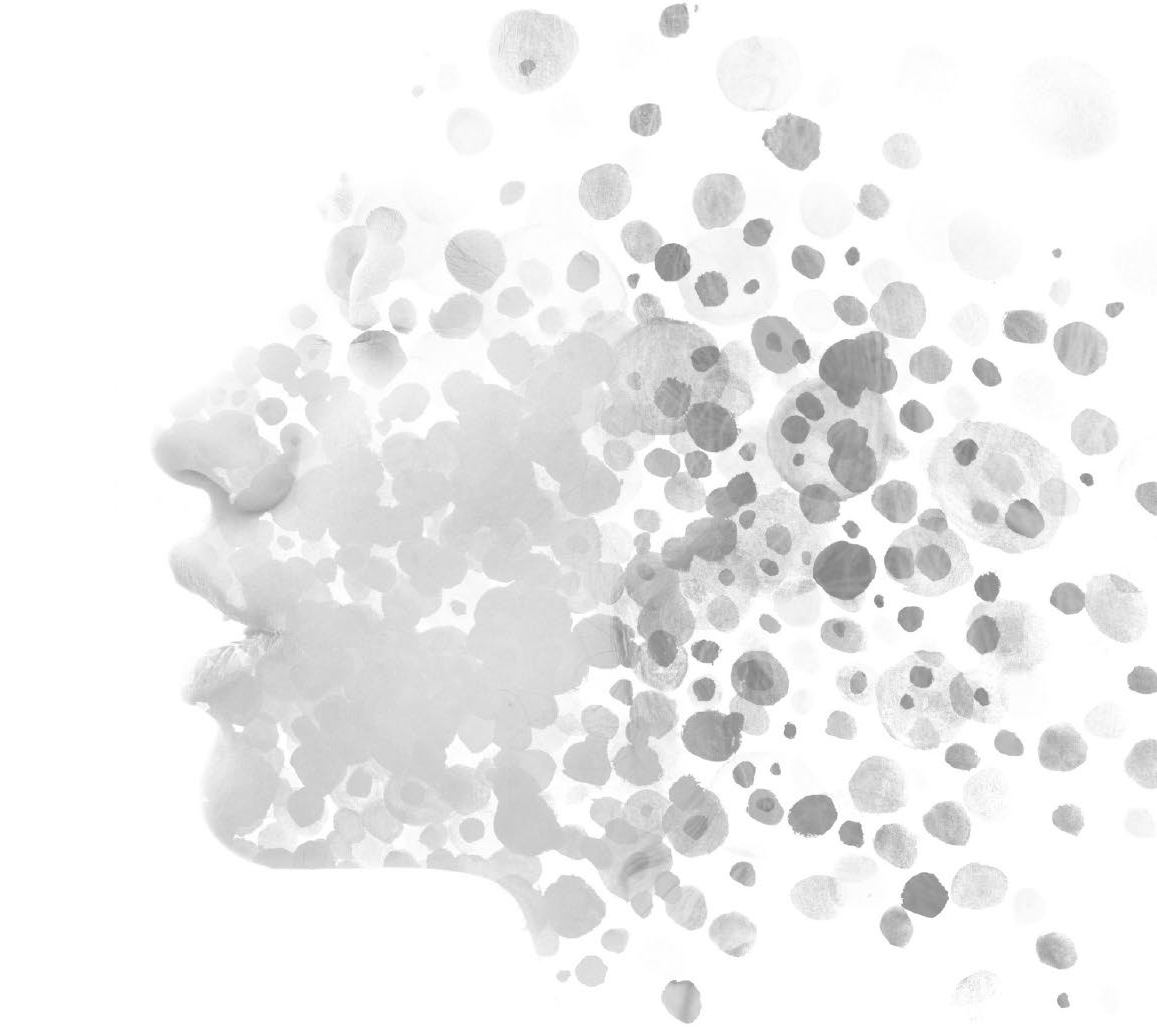
60. 改编自Sean Aday，《新闻的框架设定效应：倡导性框架与客观主义框架的实验检验》，《新闻与大众传播季刊》83卷4期（2006年冬季）：767-784页。

61. Pazit Ben-Nun Bloom，《厌恶、伤害和政治道德》，《政治心理学》35卷4期（2014年）：500页。

62. An-Sofie Claeys和Verolien Cauberghe，《保持控制：组织发言人在危机时期非言语权力表达的重要性》，《传播学杂志》64卷（2014年）：1162页。

63. Campbell和Stanley，《实验和准实验研究设计》，35页。

64. Robert F. Lorch等，《利用有效和无效的实验设计教授高、低成就课堂中的变量控制策略》，《教育心理学杂志》106卷1期（2014年2月）：18-35页。

For Review-No Commercial Use(2023)

4

# 实验类型

*真相与结论无关，而与方法论有关。*

*- 斯蒂芬·莫林纽*

• 总结使用Campbell和Stanley的分类法的不同实验类型。

• 推荐何时使用三种真实实验设计。

• 评价带有前测设计的优缺点。

• 描述准实验、自然实验和现场实验。

• 解释准实验、自然实验和现场实验彼此之间以及与真实实验的区别。

学习目标

For Review-No Commercial Use(2023)

在进行文献综述的基础上，至少在草稿形式中，正如前一章所讨论的，下一步是决定要进行的基本实验类型，即本章开头引用中所指的实验的方法或系统。在这个领域，有一本权威的参考书是Donald T. Campbell和Julian C. Stanley的经典著作《实验与准实验研究设计》。

i通常情况下，实验设计并不是以如此线性的方式进行。通常，研究人员会同时考虑本书中涉及的所有问题。

**89**

这部作品不包括参考文献只有71页。它最初是1963年的研究手册中的一章，但因为非常受欢迎，所以后来以单行本的形式继续出版。书中描述了16种不同的实验方式，至今仍然有效，尽管有些设计比其他设计更受欢迎。该书是为教育领域撰写的，因此示例都来自教育研究领域。接下来是对Campbell和Stanley的实验设计分类的总结，重点介绍社会科学中流行的六种设计以及对每种设计的评价。今天实际进行的大多数实验都是这里描述的设计的微小变体。除了Campbell和Stanley的分类中的这些设计，本章还将简要介绍准实验、自然实验和现场实验。

#### 坎贝尔和斯坦利的实验分类法

坎贝尔和斯坦利使用图形简化来描述实验，对于足球迷来说可能很熟悉，但在这里有着不同的含义。

X表示处理，通常是自变量（IV）。

O表示观察或测量的结果变量，通常是因变量（DV）。这些数据是记录下来的，可以是由研究人员记录，也可以是通过技术仪器（如心率监测器）或被试自行报告的。如果进行多次观察或记录数据，则使用下标表示，如O1、O2，表示第一次和第二次观察。

如果有一个空格（ ），没有X或O，那意味着没有给予处理或进行观察。

R表示将受试对象随机分配到不同条件的实验组。这将在第7章中进一步讨论。基本上，这是将实验区分为“真实”实验和“准实验”的要素。

前三种设计类型被归类为“前实验设计”，根据坎贝尔和斯坦利的说法，它们“几乎没有科学价值”。我在回顾的文章中发现，这些设计倾向于被报告为“探索性”的，但“前实验”是一个更准确的术语。由于缺乏随机分配，它们也可以被归类为准实验设计。在这里，我简要回顾它们，但建议使用其他描述的设计。

三种前实验设计

**单次案例研究**

有时候，这种情况在新闻报道中可以看到，当有人想将某种原因归因于某种效果时，例如，在某种干预措施后，学校的旷课率与之前的年份相比下降了。可能是某所学校实施了一种短信系统，在上午9点前提醒学生准时到校。图形表示如下：

X O

短信系统是干预措施（X），旷课率是观察指标（O）。然而，这种设计存在一些问题，其中包括观察的学生没有进行随机抽样，没有对比组的学生没有收到短信，也没有控制措施排除其他解释。例如，一些学生的父母可能开始早上做华夫饼，其他人窗外的新建筑工地吵醒了他们，有些人可能买了新车并兴奋地想要去学校展示，而其他人可能看到了有关大学重要性的演讲，并突然觉得自己需要去上学。所有这些因素（约束条件未找到）都可能解释旷课率的降低，并不仅仅是短信系统的作用。由于这些问题，单次案例研究不是一个适合进行真正社会科学实验的良好设计。

**单组前测-后测设计**

略好于单次案例研究的是在给予处理前增加一个观察或测量，称为前测。在介入干预后进行相同的观察或测量（后测），以便研究者可以将任何变化与基准测量进行比较。

O1 X O2

例如，研究人员可能在前测中测量学生的数学水平，然后进行干预，如计算机辅助辅导，然后再次测试学生，以查看他们学到了多少。前测和后测分数之间的差异代表了效应。然而，有时添加前测并不一定会使情况变得更好，因为人们倾向于在一次考试后学会如何更好地应对考试，或者当他们意识到自己正在被观察时。例如，那些多次参加智商测试的人被证明在第二次尝试时变得更聪明。在研究中，这种测试的效应被称为“测-测增益”。还可能存在其他解释。反应性是指测量某物会改变它；仅仅知道自己被观察，人们倾向于表现更好（详见第4.1个信息框）。



### 关于...盒子4.1

霍桑效应

Western Electric Company

伊利诺伊州霍桑的西部电器公司工厂，1925年

霍桑效应是一种特定类型的反应性，即人们对被观察并改变行为的反应。这在研究中很重要，因为有些人在参与研究时不会给出真实的回答。这种效应不一定是故意的；有时，人们会在不知不觉中发生变化。霍桑效应是一个流行的、尽管有人认为不适当的名称，它指的是研究对象因为知道自己正在被观察而改变行为的需求特征。它起源于霍桑工厂的工人生产力研究，该工厂是西部电器公司在芝加哥外的一个工厂，从1924年到1932年进行研究。这些研究也被称为照明研究，因为测试了不同的照明水平对工人生产力的影响；实际上，还研究了许多其他变量，包括工作时间、休息时间、地板和工作站的清洁度等。研究结果的简化版本是，当研究人员改变了工人的照明和休息时间时，他们的生产力有所提高。尽管改变保持不变，但当研究结束时，工人的生产力下降了。最初的结论是，关注工人会导致更高的生产力，但后来被重新解释为人们在被观察时会改变自己的行为。并不是所有的研究都显示出霍桑效应，研究至今仍在探讨它，特别是在健康科学、人际关系和组织行为等领域。霍桑研究还在实验设计中做出了另一个重要贡献，它显示了操作性检查的重要性，操作性检查将在第9章中介绍。在霍桑研究中，影响并不是那么重要，而是工人对这些影响的解释，理解受试者的解释是操作性检查的目的。

For Review-No Commercial Use(2023)

类似的现象被称为需求效应或需求特征，即某些人改变他们的答案或行为以取悦实验者的想法。在研究实验情境时，马丁·奥恩发现研究对象试图猜测目的并根据自己的解释改变行为。类似的现象还包括光环效应或社会期望效应，即研究参与者试图展示自己的积极形象。霍桑研究还在实验设计中做出了另一个重要贡献，它显示了操作性检查的重要性，操作性检查将在第9章中介绍。在霍桑研究中，影响并不是那么重要，而是工人对这些影响的解释，理解受试者的解释是操作性检查的目的。

如果观察者记录数据，还可能出现测量标准的变化。还有一种现象被称为回归平均，即那些得分非常高或非常低的人在下一次测试中趋向于回到中间水平。此外，如果第一次观察和第二次观察之间有时间间隔，可能发生其他事情。例如，如果记录的观察是一个人对飞行的恐惧程度，而在O1和O2之间发生了一起飞机失事事件并被报道在新闻中，这可能会改变结果。

**静态组比较**

在这种设计中，添加了一个未接受处理的第二组（O2）。

X O1

O2

例如，比较观看总统候选人辩论的人与未观看的人，或者比较接受反欺凌培训的学生与未接受培训的学生。但是这种设计也存在问题，包括无法确定两组人在重要个体特征上是否相同，例如，也许在观看辩论的群体中，共和党人比未观看的群体多，或者一组中的学生比另一组中更具攻击性。这将在第7章中的随机分配一章中详细讨论。人们的政治认同和学生固有的攻击倾向可能会影响结果。

由于我不推荐这三种设计，所以不会详细讨论它们的优缺点。相反，本章将重点讨论下面三种真实实验设计，这些设计是我推荐的。

三种真实实验设计

前测-后测对照组设计

这个设计是社会科学中最常用的设计之一。它增加了一个关键的特征，即将受试者随机分配到治疗组或对照组，从而确保两组在可能导致任何变化的重要特征上是相等的。

R O1 X O2

R O3 O4

这个设计的缺点与单组前后测设计相同，即被观察或测量两次可能会导致受试者的表现、态度或其他结果发生变化。由于这种威胁被证明是如此普遍，前测的流行度逐渐下降，实际上在真正的实验设计中，前测并非必不可少的。19

**所罗门四组设计**

常常将所罗门四组设计描述为黄金标准。它消除了前面所描述的所有缺点，此外，研究人员可以通过设立没有前测的组来判断是否存在测试效应。

|  |  |  |  |
| --- | --- | --- | --- |
| R | O1 | X | O2 |
| R | O3 |  | O4 |
| R |  | X | O5 |
| R |  |  | O6 |

但是由于成本较高、复杂且相对困难，相对而言使用它的研究并不多。20 Bausell甚至称其为“浪费”。21 在这种设计中，有四组，受试者被随机分配到所有组中。有两个对照组，其中受试者不接受处理或干预，它们作为与接受处理组进行比较的基准。一个处理组进行前测，另一个处理组不进行前测；一个对照组进行前测，另一个对照组不进行前测。这使得研究人员不仅可以比较处理前后的差异，还可以看到前测是否影响了结果。这提供了Levy和Ellis所称的“对大多数对立假设的有力回应”。22 然而，这也需要更多的受试者参与实验，而这种成本有时超过了好处。（参见研究重点4.2，了解所罗门四组设计的一个例子。）

**后测唯一对照组设计**

在Campbell和Stanley的实验设计分类中的最后一个设计是不包括前测的设计，这是我最常使用的设计，也是目前最为推荐的设计。23

R X O

R O

### 研究焦点 4.2



一项使用Solomon四组设计的研究

SAGE期刊文章：study.sagepub.com/coleman

Genç, M. 2016. “六年级学生合作学习过程评估。”《教育研究》95(1)：19-32。

这项研究采用了所罗门设计来评估特定教学策略对六年级学生科学知识的影响。这种教学策略被称为合作学习，教师将学生组织成小组，在小组内独立进行研究，共同学习信息和解决问题。合作学习的教室被视为处理组或实验组；而以传统方式向学生传授信息的教室则被用作对照组。

在这个所罗门四组设计中，有两组学生接受了合作学习干预，另外两组没有接受干预，成为对照组。其中一组对照组和一组干预组进行了前测；另外两组没有进行前测。作者称所罗门设计的使用是为了确定前测是否产生了影响，因为学生可能仅仅通过两次测试来学习。学生被随机分配到了这四个组中。下面是对所罗门设计的描述：

“所罗门四组设计是为了消除前测的可能影响而采用的。它涉及将受试者随机分配到四个组中，其中两个组进行了前测，而另外两个组没有。一个经过前测的组和一个未经前测的组接受了实验处理。尽管每个组都经历了实验后评估，但前实验评估仅在两个组中进行，一个是实验组，另一个是对照组。前两个组被视为前测-后测对照组设计，而另外两个组被视为后测-对照组设计”（第22页）。

这项研究并没有详细讨论前测是否影响了结果；然而，从平均分可以看出，接受前测的组在后测中的得分要高于未接受前测的组。然而，该研究没有报告任何显著性检验结果来确定仅前测是否有显著影响。接受前测的学生在后测中的平均分为27.70（实验组）和22.50（对照组），而未接受前测的学生在后测中的平均分为26.74（实验组）和21.91（对照组）。为了说明设计并便于比较前测组和非前测组的后测得分，我重新制作了作者的表格，以下是重新制作的表格：

需要注意的是，前测和后测使用了不同形式的测试，而不是相同的测试，这有助于防止通过两次测试来学习。

合作学习处理有效，合作学习教室中接受前测和未接受前测的学生在实验结束时的科学测试中表现明显更好。

一些社会科学学科比其他学科更常使用所罗门设计，因此了解同事对此的看法非常重要。

（未完）

所罗门四组设计实验的均值和标准差

|  |  |  |  |
| --- | --- | --- | --- |
| **组** | **预测** | **指令类型** | **后测得分 M（标准差）** |
| 处理组 1 | X | 合作学习 | 27.70 (2.531) |
| |  |  |  |  | | --- | --- | --- | --- | | 处理组 2 |  | 合作学习 | 26.74 (1.797) | |  |  |  |
| 对照组 1 | X | 常规课程 | 22.50 (2.219) |

|  |  |  |  |
| --- | --- | --- | --- |
| 对照组 2 |  | 常规课程 | 21.91 (2.327) |

在一项比较前测-后测设计和仅后测设计的研究中，Gorard（引文24）发现仅后测设计的结果“更少误导性”。即使是Campbell和Stanley（引文25）也表示前测在真实实验设计中并非必需，并解释了人们对前测的误解。当前测的目的是确保被分配到不同条件的受试者之间的等价性时，随机分配是一种足够的预防措施，因此不需要前测。他们指出，几乎所有罗纳德·费舍尔（Ronald Fisher）的农业实验都没有前测。

前测仍然被某些领域视为重要；这是在该领域中的标准。前测与处理之间的交互作用在大多数情况下并不大，但特别是在教育、心理学和社会学等领域，效应较大，不应忽视。一些研究则处于前后测的界限，例如进行了使用前测的试点研究，如果排除了测试效应，那么进行实际实验时就不再进行前测（引文27）。 （参见"如何做"4.3框中有关如何描述仅后测对照组设计的示例。）

这些设计代表了一种基本结构；实际的实验可能会有细微的变化，但仍然是可接受的。到目前为止，本章回顾了三种被归类为“真实”实验和三种被归类为“准实验”的实验设计，根据Campbell和Stanley的分类法。Campbell和Stanley还描述了其他十种被称为“准实验”的设计。本书不会详细介绍它们，而是总结准实验的基本特征，并将读者引向Campbell和Stanley的书籍以获取更多详细信息。

#### 准实验研究

真实实验和准实验之间的关键区别在于，在准实验中，受试者没有被随机分配到条件中，群体也不一定是