

北京大学
心理学教材

心理实验 研究基础

朱 滢 主编



北京大学出版社
PEKING UNIVERSITY PRESS

北京大学心理学教材

心理实验研究基础

朱 滢 主编



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

心理实验研究基础/朱滢主编. —北京:北京大学出版社, 2006. 10

(北京大学心理学教材)

ISBN 7-301-10642-4

I. 心… II. 朱… III. 实验心理学-研究生-教材 IV. B84

中国版本图书馆 CIP 数据核字(2006)第 113615 号

书 名: 心理实验研究基础

著作责任者: 朱 滢 主编

责任编辑: 陈小红

标准书号: ISBN 7-301-10642-4/C · 0406

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn> **电子邮箱:** zpup@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62752038

出版部 62754962

印 刷 者: 北京大学印刷厂

经 销 者: 新华书店

787 毫米×960 毫米 16 开本 11.75 印张 237 千字

2006 年 10 月第 1 版 2006 年 10 月第 1 次印刷

定 价: 20.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子邮箱:fd@pup.pku.edu.cn

本书献给我的哥哥朱荣和嫂嫂潘静致
——他们指引我走进研究生的队伍

作者序

1989年至今我为北京大学心理学系的研究生讲授实验方法课,这本书就是在讲课的基础上编写的。任何一项心理实验都必须明确地或内隐地面对一系列的问题。这些问题包括:概念的定义、实验类型、实验研究中的样本、测量、实验设计与统计,等等。本书在占有各心理学领域丰富的文献基础上,分析这些问题并给出一些可能的解决办法。

传统的显著性检验方法长期以来受到严厉批评,本书对此给予足够的重视,设有一章讨论显著性检验,并在另一章介绍元分析的方法,因为元分析方法被看成是克服显著性检验基本缺陷的有效途径,在元分析一章中还介绍了元分析的最新进展,即脑成像研究中的元分析方法。实验研究越来越多地应用于临床、商业等领域,但由于在这些领域不能对相关因素作精细地控制,因而不能套用实验室的设计方法,本书有一章介绍准实验设计来解决此类问题。实验中的因果关系是实验设计的基础,减小实验误差是实验设计的目的,本书有两章分别讨论因果关系和实验误差。目前心理学实验研究中大量地使用计算机,不仅仅用它来处理数据(特别是脑成像结果的海量数据),而且用于控制实验、对所研究的问题提供模型,等等,本书有一章讨论计算机在心理学研究中的应用。

20世纪80年代以来,心理学与脑科学的结合导致了心理学发展的新阶段即认知神经科学,脑成像技术是认知神经科学强有力的工具,本书对fMRI作了专门的介绍,包括基本原理,fMRI实验(实验设计、数据获取、数据分析)以及在注意、情绪、记忆、社会认知神经科学等领域fMRI的研究实例。

本书较多地引用了Plutchik(1983)的论述^①,我认为Plutchik的书是讨论心理实验的经典著作,许多论述和思想至今仍有启发意义。该书是已故教授陈舒永先生为研究生课程指定的基本教材,我一直在教学中使用。

我邀请了范津(Jin Fan)博士编写“功能磁共振成像”一章,他现在纽约Mount Sinai School of Medicine任教;王洪斌(Hongbin Wang)博士编写“计算机与心理学研究”一章,他现在The University of Texas任教;郭春彦、张力博士编写“元分析”一章,他们现在首都师范大学任教;卓彦研究员与傅小兰研究员长期以来为研究生的课程作专题

① Plutchik R (1983). Foundations of Experimental Research. New York: Harper & Row

报告,但因为工作繁忙未参与本书编写。我对他们的无私帮助深表感激。

北京大学出版社的编辑陈小红为本书的出版付出了辛勤的劳动,王阿尼等人耐心地打印全书文稿,张力和隋洁帮助整理文稿,我的妻子汪慧丽为我长期的工作和写作创造了一个优良的环境,我向他们表示衷心的感谢。

朱 滢
于北京大学燕北园
2006年3月

写给读者

本书的写作目的不是对各领域的研究加以总结,而是讨论怎样才能做好心理实验研究。在此我向读者推荐 Plutchik (1983) 指出的正确学习途径: “It seems to me that a student can develop an appreciation for research only by going to original experimental reports published in contemporary journals.” 请注意 “only” 这个词! 也请注意 “original experimental reports” 几个词, 最优先阅读的是原始的实验报告, 然后才是一般理论文章或综述。

除了要阅读各专业方向的杂志之外, 有意于从事心理学研究的读者应该优先阅读下列杂志, 因为刊登在这些杂志上的心理学文章水平最高。这些杂志是: Science, Nature, Trends in Cognitive Science (几乎没有实验报告), Proceedings of the National Academy of Science USA, Psychological Science, Current Directions in Psychological Science, American Psychologist (几乎没有实验报告)。

目 录

第 1 章 引论	(1)
一、实验金字塔	(1)
二、实验中的各种决策过程	(4)
三、记忆实验的四面体模型	(6)
第 2 章 概念的定义	(10)
一、为什么需要新的概念	(10)
二、心理学家是如何定义概念的	(12)
三、概念的定义与科学研究的关系	(18)
四、面对科学难题,我们需要新的科学概念	(19)
第 3 章 实验类型	(22)
一、二价实验	(22)
二、多价实验	(24)
三、参数实验	(25)
四、从二价实验到参数实验	(27)
五、双重分离实验	(28)
第 4 章 实验研究中的取样	(33)
一、全体与样本	(33)
二、被试样本	(34)
三、物种样本	(38)
四、刺激样本	(39)
五、实验条件样本	(41)
六、测量样本	(42)

七、不同样本之间的差别	(43)
第5章 统计显著性检验	(46)
一、统计显著性检验的逻辑	(46)
二、零假设的逻辑	(47)
三、实验效果大小的测量	(49)
四、显著性检验的基本缺陷	(52)
五、评价实验研究的基本原则	(55)
六、增强统计检验能力的一些方法	(55)
第6章 元分析方法	(59)
一、元分析方法简介	(59)
二、元分析方法的应用	(67)
三、元分析方法与传统统计方法的比较	(73)
四、元分析方法的展望	(83)
第7章 实验中因果关系的概念	(86)
一、因果关系的一般概念	(86)
二、米勒确立因果关系的方法	(88)
三、脑成像研究中的推断方法	(91)
第8章 误差的来源与控制的性质	(92)
一、误差的各种原因或来源	(93)
二、误差是一把尺子	(96)
三、实验研究中控制的概念	(99)
四、单个实验研究的局限性	(99)
附录 黄禹锡事件给亚洲的科学教训	(100)
第9章 实验设计方法	(102)
一、随机组设计	(102)
二、组内设计	(104)
三、随机区组设计	(107)
四、拉丁方设计	(108)

五、被试样本的大小问题	(109)
六、使用一个被试的实验有意义吗	(110)
附录 过分信赖权威及其对实验设计的影响	(112)
第 10 章 准实验设计	(119)
一、准实验设计	(119)
二、效度问题	(124)
第 11 章 测量的性质	(128)
一、测量的一般概念	(128)
二、心理学中的测量	(129)
三、经典物理学与心理学在测量上的比较	(130)
第 12 章 计算机与心理学研究	(135)
一、计算机作为心理学研究的工具	(136)
二、计算机作为心理学研究的理论和方法论	(144)
第 13 章 功能磁共振成像	(160)
一、功能磁共振成像的基本原理及特性	(160)
二、功能磁共振成像的应用	(170)

1

引 论

着手进行一项心理实验研究,它的全过程到底是怎样的?包括哪些内容?有哪些基本问题需要研究者思考并给予适当解决?下面我们从实验金字塔、实验中的各种决策过程以及记忆实验的四面体模型等三个不同的方面,对上述问题进行讨论。

一、实验金字塔

Anderson(2001)用一个实验金字塔(见图 1.1)来表示,当我们进行一项实验研究时要解决哪些问题才能实现研究的目的。这些问题包括顶层的统计推论和实验设计到底层的概念框架。愈底层的部分愈是基本的;较低层次的效度是较高层次效度的前提。各个层次之间并不截然分离如像图中虚线所暗示的那样,它们彼此连接构成一个有机的整体。

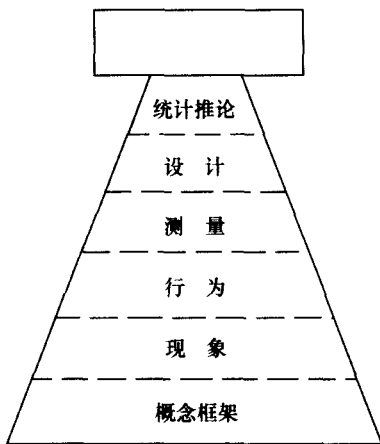


图 1.1 实验金字塔(Anderson, 2001)

简单说来,统计推论决定实验结果是否可靠;实验设计将要研究的问题安排好

或形式化;测量就是因变量;行为是测量的对象;现象是所测量的东西在心理学上叫什么;概念框架是研究的出发点与归宿;我们为什么要进行一项实验研究,怎样解释获得的结果。下面我们分别简要地介绍这些问题,以便大家形成一项实验研究的总体印象,在后来的各章中我们将详细讨论各个问题。

(一) 统计推论

相对而言,显著性检验(significance test)在实验金字塔的框架内或在科学研究中是一个较不重要的问题。当然,我们需要对实验结果进行显著性检验,以便确定该结果是确实的而不是由机遇产生的。但是,显著性检验的前提是恰当的实验设计。例如,一项研究表明,新疫苗导致 16% 白喉症病人死亡,而原有的药物治疗只有 8% 的死亡率,两者达到显著性差别。表面看来新疫苗确实无效,但仔细分析后发现,死亡率为 16% 的病人多为晚期重症病人,而 8% 死亡率的病人多为轻度症状病人。由此可见,统计作为一个研究工具不可能保证实验结果的科学意义,就像计算机作为一个工具不能保证计算结果的科学价值一样。另外,也不要死板地去应用显著性检验。例如,两项类似的实验发现显著性水平分别在 0.04 与 0.06,它们实际上具有良好的一致性,虽然一项结果达到了显著性水平而另一项没有达到。又例如,一项超感觉知觉(extrasensory perception)实验达到了 0.05 的显著水平,但以往的有关研究从未达到 0.05 的水平,那么该研究结果值得怀疑。

实验研究结果的证实不仅仅通过显著性检验来实现。不同的研究者在各自的实验室或同一研究者在后来的其他场合能重复该结果,更是该结果真实性,即具有良好的信度的证明。Plutchik(1983)说得好,借用一句谚语:“一次重复抵得上一千次 t 检验。”

在脑成像研究中,只要有关脑区激活达到了统计显著性水平,而行为数据由于被试人数较少(如 Craik 等人(1999)在 PET 实验中使用 8 名被试;Zhang Li 等人(2005)在功能磁共振成像实验中使用 7 名被试),未达到显著性水平是可以理解的。

(二) 实验设计

实验设计是实验研究的蓝图,反映了研究问题怎样组织、安排,怎样转化为一个实验程序,没有设计就无法进行实验。有人认为实验设计是统计学的一部分内容(张厚粲,1990),也有人认为实验设计一定要建立在统计学原理的基础上(舒华,1994)。总之,二者有密切的关系。例如,设计中要控制好额外变量(extraneous variable)以防止混淆,一个基本的控制方法就是随机分配,而随机分配是统计学的重要内容;现代心理学实验的研究涉及多个变量,因此实验设计通常包括多个变

量,像因素设计,所以,分析多个变量之间的交互作用十分重要,而交互作用是现代统计学的新进展之一。

有了恰当的设计,研究问题的意义还依赖于测量,实验到底测量了什么。

(三) 测量

测量把行为的世界与科学的世界连接起来。换句话说,测量把客体与事件的客观世界转换(transformation)成概念与符号的主观世界。例如,你学习了英文单词“observation”,如果用自由回忆的方式测量你的学习,得到的结果是关于你的外显记忆(explicit memory);如果用补笔的方式(word-fragment completion)(例如,把“o_se_vat_on”填成一个有意义的单词)测量,得到的结果是关于内隐记忆(implicit memory)的。通过测量,我们把客观的学习事件转化为记忆这种主观过程。在实验研究中测量是由实验任务、仪器和实验程序决定的。效度与信度是测量中两个最重要的概念。效度(validity)指测量到的结果的确是由设计中的自变量引起的,而信度(reliability)指测量结果的一致性,即能否重复,两者都依赖于实验金字塔中最下面的三层:行为、现象与概念框架。

(四) 行为

选择什么行为作为研究对象是研究者的实验设置决定的,实验设置包括:物种、任务、仪器、实验程序和反应测量,等等。这些选择同时还决定了测量数据的含义。例如巴甫洛夫(Pavlov)对狗的唾液反射作用进行了系统的研究,建立了条件反射学说。唾液分泌的行为易于测量,同时又有普通意义,即代表着或测量着许多心理现象(学习、辨认、神经症、睡眠、神经过程的兴奋与抑制等)。又例如,“最后通牒游戏”(Ultimatum Game)被广泛用于决策过程的研究。

(五) 现象

我们的研究目的通常是某种现象——记忆、颜色知觉、语言、社会态度等。而实际上,我们研究的是某种观察到的行为,我们认为这些行为是对现象的可靠与有效的测量。这就是行为的事实与命名为某种现象之间的差别,涉及对行为在概念上的解释。有时候观察到的行为不一定是我们想要研究的现象,这就是实验研究中的混淆,即自变量与额外变量或无关变量的混淆。

行为与现象的差别还涉及普遍规律的概括问题。艾宾豪斯想要研究人的记忆规律,他使用无意义音节作为学习材料,得到的一些结果有普遍意义,如遗忘曲线,但其他一些结果不具有普遍性,如背诵次数多导致优良记忆。这一结论忽略了学习策略、加工水平对学习的影响,因而是片面的。

(六) 概念框架

概念框架的作用表现在两方面。第一,选择什么问题进行研究。不同的问题大到不同的心理学流派方向上的问题,如行为主义的、认知心理学的、认知神经科学的或社会认知神经科学的,小到具体不同的问题,如记忆的提取、记忆抑制与记忆错觉。例如,认知神经科学兴起之后,人们用心理事件与脑功能的概念取代心理与计算机程序之间的类比,而大脑具有兴奋与抑制过程(计算机并没有抑制过程),因而引起了人们对记忆的抑制过程研究的兴趣(Smith, et al, 2000)。一般记忆研究强调记忆的准确性,但把记忆看做是重构过程的观点,使记忆错觉的现象受到重视,在记忆错觉的研究中人们测量“记住”未呈现过的单字的数量(张力,1998)。

第二,概念框架的不同,导致对实验结果的理论解释不同。实验金字塔的帽子代表着对获得的实验结果的分析与解释,通常就是在实验研究文章中的讨论部分,它涉及金字塔的各个层次。对实验结果的解释本质上是超统计的推论(extrastatistical inference)。例如,面对记忆中的实验性分离现象(experimental dissociation),Tulving 等人(1990)提出多重记忆系统的思想给予解释,而 Roediger 等人(1989)则提出记忆过程的程序观点加以解释;前者属于认知神经科学,后者属于认知心理学。

每个研究者的概念框架、知识体系都是不同的,也时时在变化发展中,这是一件好事。这意味着不同的研究者将研究不同的现象,甚至研究同一现象的不同方面,并提出不同的观点来进行解释。这使得研究大大地丰富起来,充满活力,最终推动科学进步。

二、实验中的各种决策过程

(一) 什么是心理学实验?

心理学是实验科学,那么,什么是心理学实验呢? Simon(1974)给出了一个定义(或说明):心理学实验要求额外变量(extraneous variable)保持恒定,而仅仅操纵自变量去影响因变量。并且他还设定一个零假设:因变量的平均值在不同的实验条件下没有差异,如果所获得的实验数据否定零假设,那么,实验者就得到了一个可靠的结论,即一点有用的信息——因变量是明显地受自变量影响的。

迄今为止,心理学知识的积累基本上都是在这个框架内实现的。Simon 的定义叙述了心理学实验的基本特点,即实验中的额外变量要保持恒定以及实验结果的显著性检验。但这个定义过于强调了显著性检验在实验研究中的作用,关于这

个问题我们将在第5章进行讨论。

(二) 进行心理学实验的四个理由

Plutchik(1983)指出,所有实验研究追求的目的是增进我们对所研究事件的理解以及控制与预测事件的能力,但除此之外,还有一些具体的理由说明我们为什么要进行特定的实验研究。

1. 决定两个或多个变量之间的关系

有一些探索性实验,研究者想要确定在哪些条件下某种事件将会发生。假设要比较甲、乙两种声音刺激的可听度,而且这两种声音刺激又是发生在两种不同噪声背景上(噪声一和噪声二),每种噪声又各有两个强度水平(90和120分贝),这样,声音刺激(甲与乙)、噪声(一与二)、噪声的强度(90分贝与120分贝)这些变量就成了影响可听度的三个变量。系统地收集这类资料是实验的基本功能,它成为心理科学的基础。一个心理学理论得以建立或发展,必须依赖实验所提供的已证实的经验发现。

2. 扩展变量研究的范围

由于任何一项实验都是在特定的即有限的条件下进行的,因此,完成一个实验之后总会提出新的问题来继续探索。例如,自我参照效应(self reference effect)的研究表明,西方被试的自我参照记忆效果远远优于母亲参照(Klein, et al, 1989)。但进一步的研究却得出了不同的结果,中国被试的自我参照记忆效果并不优于母亲参照,而是处于同等水平(Zhu, 2002)。类似研究的积累,使研究者在实验范式上归纳出一个重要的跨文化心理学研究的规则:类似的认知任务由不同文化背景的个体来完成时,其过程是不类似的(Li, 2003)。就上面的例子来说就是,中国人与西方人在完成同一任务(母亲参照)时,他们各自的认知过程是不同的,这导致了不同的记忆效果。

3. 增强所报告的发现的信度

这涉及对已有发现的重复(replication),或者在原有条件下重复,或者在扩展的条件下重复。这一点在任何科学实验中都是必要的。重复是实验的最重要的功能,因为一个人不可能在实验研究之前对所有有关实验的因素都能采取预防措施来避免误差。如果不同的实验者使用不同样本的被试而能确证已有的发现,我们对该发现的信心就会大大增强。心理学不少发现不能重复,因此实验的这一功能值得强调。Gauch(2006)还指出,科学家喜欢用两次重复的平均数代替一次未经重复的结果,因为平均数好像更准确些。但是,实际上这不意味两次重复一定更准确,仅仅由于机遇第一次重复可能很接近真实的结果,而第二次重复就可能远离真实结果,两者平均实际上不如一次未经重复的结果。按照统计学的规律,重复5次

的平均结果有 73.2% 的把握 (confidence of success), 这比一次测量的结果更准确。要增加重复成功的比率到 90%, 需重复 40 次。Gauch 的意思是, 一项实验发现必须经过多次重复才能为科学界所接受。

4. 检验理论

进行实验的第四个理由是检验理论。例如, 我国心理学家 Chen Lin (2005) 提出了拓扑性质知觉理论, 他和他的学生二十多年来进行的实验研究都是围绕拓扑性质知觉理论开展的, 这一理论的重要性已开始为西方学者承认。

Plutchik (1983) 指出, 对各种心理学理论, 假设的检验通常得不到清晰明确的结果, 因为理论是概念体系, 没有决定性的实验可以证明或否定它。这种情况并不奇怪, 因为大多数理论的价值在于启发出相关的研究而不是预期新的事实。从这个观点看来, 一个理论只能是被另一个更好的理论推翻, 而绝不会仅仅是由一些相反的事实所推翻。

(三) 实验中的各种决策过程

研究者在计划实验、进行实验、数据处理等包括一系列步骤的一项实验研究中, 总会有意识地或内隐地考虑如下的问题并作出相应的决定: 概念的选择、样本、实验类型、实验设计、测量、统计与数学方法的使用、结论的抽取与解释, 等等。

三、记忆实验的四面体模型

Roediger 和 Meade (2002) 认为, Jenkins 的记忆实验的四面体模型 (见图 1.2) 提供了一种很好的方式来思考人类记忆的实验。

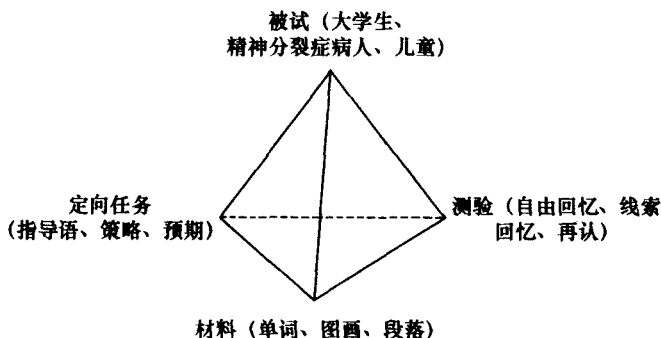


图 1.2 记忆实验的四面体模型 (Roediger & Meade, 2002)

首先, 记忆研究者进行实验时必须四个维度上进行选择, 无论哪个特定维度是研究者感兴趣的。这四个维度是: (1) 被试的类型; (2) 实验中使用的材料;

(3) 定向任务(orienting task)(或被试受测查的情景特征);(4) 测量记忆的测验类型。

其次,Jenkins 的四面体模型还说明,任何实验都只创建出某些有限的变量或因素,并且该实验还是在其他的变量或因素保持恒定的情况下进行的,因此每一项实验结果都是镶嵌在保持恒定的其他因素组成的背景之中的。不过研究者常常隐约地假定,他们在特定有限的实验条件下获得的结果将概括化地应用于其他条件。

第三,四面体模型提供了一个有用的方式来考察记忆的主要发现。例如,在记忆的测验类型方面,用内隐记忆测查实验结果与用外显记忆测验的结果十分不同。一些脑损伤病人丧失了大部分记忆新信息的能力,可称其为顺行性遗忘症,此类病人所有的外显记忆测验的成绩都很差。然而,在 20 世纪 60 年代,英国心理学家 E. Warrington 和 L. Weiskrantz 却发现这种病人在内隐记忆测验中表现出了极佳的正常状态,现在这一发现已被重复了很多次。又例如,当语言中其他特征(如词类、长度等)保持恒定时,高频词比低频词更容易被回忆起来,然而,当我们采用再认测量时,又变成低频词比高频词再认得更好!这就是文献中所述的频率作用的二重性(frequency paradox)。上述两个例子表明,有很多测量记忆的方法揭示了记忆中不同的过程和系统。

上述记忆实验的四面体模型原则上也适用于心理学的其他领域,因此我们可以预期,在任何一项实验中如果研究者能在被试、材料、定向任务和测验这四个维度中的任何一个有所创新,那实验结果肯定会有新意。

在结束本章之前,让我们来看一看图 1.3。这是 Gauch(2005)为科学的方法论所作的图解,他认为科学方法的普遍原理适用于整个科学事业(心理学是其五个具有代表性的学科之一),而专门的技术仅限于特定的学科和分支学科。科学家如果成功地普遍原理与专门技术结合,就会既加强他们的成果产出率,又拓宽他们的视野。

Gauch(2005)反驳了不存在科学方法的普遍原理的说法。他说:“其实,只要问三五个具体的问题,就足以消除疑惑、廓清正误,我们不禁要问:天文学家肯定要运用演绎逻辑,难道微生物学家根本就不使用演绎逻辑吗?心理学家经常使用归纳逻辑以及统计方法,从大量数据中抽取出必要的结论,难道地质学家干脆就不用这些方法吗?在生物学研究中经常运用概率论和计算方法,社会学就一点不使用这些方法吗?医学研究工作者十分关注简约模型的运用,利用它解释问题,作出判断,难道电气工程师就不需要简约模型吗?物理学总是预先假定物质世界是客观存在的并且是可以理解的,难道遗传学就会出现例外情况吗?”Gauch 还指出了科学方法普遍原理的核心部分,即图 1.3 中心的重叠部分,它包括假说的形成和检测、演绎逻辑和归纳逻辑、简约性以及科学的预设、科学的有效作用范围和科学的