

重 审 实 验*

方 文

本文以内在效度——外在效度为维度,构造了一种社会心理研究方法的评价框架,并且分析了社会心理研究中的反应性;在此基础上,集中研究了基本的实验逻辑,以及对实验研究的种种批评;结论在于:实验在社会心理研究方法体系中占有独特的地位,而对它的任何评价应抱有一种审慎的态度。

作者:方文,男,1966年生,博士,北京大学社会学系博士后。

对秩序的追求是人类一切智慧活动的基本动力和需求。人类的一切智慧活动背后也都隐含着这样一个基本的潜在假设,或者说一种基本的潜在心理倾向:在种种杂乱无章的现象背后,通过特定的程序,我们可以剔除种种干扰着的混乱(Chaos)、无序(disorder),去发现所隐藏着的秩序(order)。这种种秩序或者是现象本身所具有的,或者是因为人的理念图式,人的先天综合判断去投射、重构现象后的结果。人类心智进步多少,混乱或无序就退却多少,而有序也就显现多少。

和自然科学一样,社会科学(包括社会心理学)的研究也基于同样的基本假设。具体说来,可以分为四点:

1. 整个世界包括人类行为及社会现象有规律、有秩序可循;
2. 这种秩序或者规律可以采用种种研究分析程序而获得;
3. 这种种研究分析程序是可教的(teachable)和可交流的(communicable);
4. 研究结果的正确与否可以在科学共同体内部相互印证,并且在实践中受到检验。^①

就社会心理学而言,这种种研究分析程序则构成社会心理学的方法体系。在这种方法体系中,实验受到最为广泛的使用。有3/4的社会心理研究采用实验法^②,并且其中有2/3的研究情景是在实验室进行的^③。

与其应用的广泛性相伴随,实验也一直受到最为广泛的争议。一方面,实验法的介入被认

* 本文是哲学博士论文《社会行为学论纲》中的一部分,导师为陈元晖先生。写作及修改过程中还得到李伯黍先生及沈德灿先生的帮助。

^① 参看:魏镛:《社会科学的性质及发展趋势》,(1973),载《云五社会科学大辞典·社会学卷》,台湾商务印书馆,第58—60页。

^② Highce, K. L., et al 1982 Social Psychology Research during the 1970s: Predominance of Experimentation and College Students. *Personality and Social Psychology Bulletin*, 8: 180—183.

^③ Adair, J. G., et al 1985 Ethical Regulations and Their Impact on Research Practice. *American Psychologist*, 40: 59—72.

为是科学社会心理学即社会分析学诞生的标志^①；另一方面，实验的过份滥用被认为是现代社会心理学危机的一个主要根源^②。实验的地位、效用及其限度到了必须予以理性的审查的时候了。

一、社会心理研究方法的评价框架

在社会心理学的研究方法体系中，作为工具的方法只为了一个目的：力求对种种社会心理及社会行为现象进行秩序重组或者规律重组，——其结果往往以理论或假说的形式来表征：或者是相关假说，或者是因果假说，并且这些假说可由他人重复验证。研究者由于其偏好，以及研究目的及选题的不同，往往采用不同的方法，如现场观察（field observation），现场实验（field experiment），问卷调查（questionnaire surveys），深度访谈（depth interview），或者实验室实验（laboratory experiment）。但无论如何，研究所采用的方法必须满足二个准则：内在效度（internal validity）及外在效度（external validity）。所谓内在效度意味着社会心理研究者采用某种方法所获得的结果是表征真实的或者有意义的差异，而不是研究程序本身的人工制品（artifacts）^③。假设 A 和 B 在两所大学的研究学生考试中作弊行为的水平，并且两者一致同意在同样的学期中采用同样人数的班级，且同意把“作弊”界定为“考试期间偷看别人的试卷”。但两人没有确定监考老师或研究者在教室中所站的位置，结果 A 坐在教室里一个角落里，而 B 则在教室后面观察。这样如果比较两所学校总体的作弊水平，前者高于后者。并且知道 B 的学校校纪较之 A 的学校为严，那么是否可以得出严格的校纪能有效地防止作弊行为呢？不能。因为这里的行为差异不一定是真实存在的，而绝大多数可能是研究中的虚假现象——因为 B 的学校的学生意识到监考老师的严格注意而使作弊行为剧减。那么这个研究结果所表征的就不是真实的社会行为现象，其内在的效度就低。内在效度表征出研究中研究者各个环节控制的连贯程度。研究中控制越好，其内在效度愈高，研究结果愈能表征真实的行为差异。

而外在效度则表征研究的概括性（generalizability），或者说研究结果能否推广到其他样组的被试，或者其他的情景中^④。即使在两所大学中获得了作弊行为与校纪松严之间严格的关系，它是否可以广推为“校纪严的学校其作弊水平比较纪松的为低”呢？或者用专业术语“较高的惩罚的威胁会阻止作弊”呢？为了适当地回答这个问题，还必须去比较更多学校之间的作弊行为的差异以及其他。用越多的方式去验证一个假设，并且结论一样，那么原初的研究的外在效度愈高，换句话说，外在效度涉及到所研究对象的代表性（representativeness）问题。如果研究对象能够表征类似情境下的其它所有对象，那么其代表性就高，以这样的对象作为样本（sample）的研究，其外在效度就愈高。在具体的操作过程中，往往是通过随机抽样（random sampling）来获取有代表性的样本。

更进一步，由于研究中的对象是具有自我意识的人类个体或群体。研究对象在研究情景中如果清晰地意识到他或他们正在作为研究对象，那么研究对象的自我知觉（self-awareness）就

① Hollander, E. P. 1981 Principles and Methods of Social Psychology (4th ed.) N. Y. : Oxford. pp10-13.

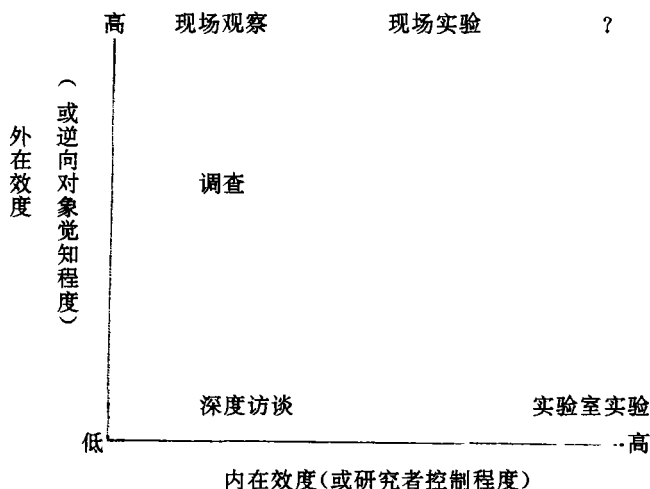
② Vitelli, R. 1988 The Crisis Issue Assessed: An Empirical Analysis. *Basic and Applied Social Psychology*, 9:301-309.

③ Shaver, K. G. 1981 Principles of Social Psychology, 2nd ed. Cambridge, Mass: Winthrop Publishers, Inc. p34.

④ *Ibid*, pp34-35.

作为具有较大影响的干扰因素渗透到研究过程中,从而减低了研究的概括性,也就是降低了研究的外在效度。因此我们可以用研究对象的自我觉知来作为外在效度的一个参照标准:外在效度跟逆向对象觉知程度正向相关。

不同的方法所具有的内在效度及外在效度各不相同。我们可以依据内在效度——外在效度,或者研究者的控制程度——逆向对象觉知程度这二个维度,作为参照系或者评价框架,来比较几种常用的社会心理研究方法,见下图。



(参考 Hollander, E. P. 1981 和 Shaver, K. J. 1981)

社会心理研究中理想的研究方法应该能同时良好地满足内在效度——外在效度这二个维度。但这样的方法现在还没有发现。对内在效度——外在效度一方的重视往往以对另一方的损失为代偿。在这个框架中,我们就会客观地依据这个评判尺度,避免用一种方法去取代或者贬低另一种方法。并且也会意识到只能在具体的研究情景中去评价方法的优劣,同时也会在具体的研究情景中去考虑替换方法或者方法组合的可能性。

二、社会心理学研究中的反应性(reactivity)

在许多情形下,作为研究对象或者被试,他们意识到自己不仅仅在行动,而且还在创造研究数据或者资料,因此以人为研究对象的社会心理研究的结果,总是受到被试确证其行为表现被记录下来以供分析的这种知识的干扰。所获得的研究资料数据受研究对象的干扰程度称之为研究技术的反应性^①。这种潜在的误差来源,几乎出现在所有的研究方法的应用中。Shaver 将其归为二组:导源于被测试的觉知的问题和导源于反应定势(response sets)问题^②。

1. 被测试的觉知

几乎在所有的研究情景中,都发生这样的情况,被试不是“自然地行动”(behaving

① Shaver, K. G. 1981 *opcit* p36.

② *Ibid.*

naturally),而是“象好被试那样行动”(behaving like a good subject)。这种成为好被试的角色扮演,以期获得研究者的积极评价,被 Rosenberg 称之为“评价的领悟”(evaluation apprehension)^①。这种对“评价的领悟”或者成为好被试的角色扮演,极大程度上干扰了研究结果。

在一个令人吃惊的实验中,Orne 设计了一些非常单调且无意义的工作,如在纸头上不停地做加法,或者不断地撕纸头。他以为一个清醒的被试会拒绝做,至少是放弃,他惊讶地发现被试干个不停,且在事后评价中,称这是件饶有兴趣的工作^②。Orne 将被试的这种反应又称之为需求特征(demand characteristics)。这种令人惊奇的结果很容易让人想到 Milgram 有关对权威服从的实验研究。

无论是“好被试的角色扮演”,或“评价的领悟”,或者“需求特征”,都与被试意识到自己在被研究这个问题连在一起,它是反应性的第一种类型。

2. 反应定势

反应定势意味着在社会行为研究中,被试不管特定的情形或特定的信息内容,普遍地以某种特定的方式作出反应。如在访谈问题(interview questions)上,他们总是同意某种陈述。这种情形可称之为默认反应定势(acquiescence response set)^③。当然这种定势在访谈情景中很容易消除,如把访谈问题一半用正面陈述,一半用反面陈述。

另外的反应定势,如社会认可性(social desirability)即被试依据特定的研究者所呈现的亚文化规范来调整自己的反应,则更难处理。如果研究者在访谈研究中雇佣一个访谈者(interviewer),在访谈开始前吹嘘自己如何偷汽车等“奇迹”,那么,在访谈过程中,被试者所描绘的自己的偷窃行为的比例肯定会提高^④。社会心理研究者正在努力采用种种措施来使反应性尽可能降低:包括精确的研究设计,以及采用档案材料、文字分析等等。

三、基本的实验逻辑

在任何标准的实验心理学或社会心理学教科书中,都有关于实验程序的详细论述。概括地说,实验程序的基本思路是控制自变量或刺激变量以观察因变量或反应变量的变化,从而确定自变量和因变量之间的因果关系。它有三个基本要素:实验者,通过基于心理统计的严格的实验设计,控制或者排除种种无关的刺激变量,而使感兴趣的刺激变量凸现出来;被试,作为社会行为的研究对象,在刺激的作用下发生种种的反应;实验情景,绝大多数是在模拟自然的社会环境的实验室中进行的。

假设我们在研究电视对儿童态度及行为的影响。通过自然情景下的观察,我们发现观看暴力电视节目(violent television programs)多的儿童比少看的儿童更多地表现攻击或侵犯性行为(aggressive behaviors)。这种观察也许会导致这样的结论:儿童从电视中学习某种行为——我们可以把它转换成社会心理学的语言:学习(自变量,如观看暴力节目)和社会行为(因变量,如攻

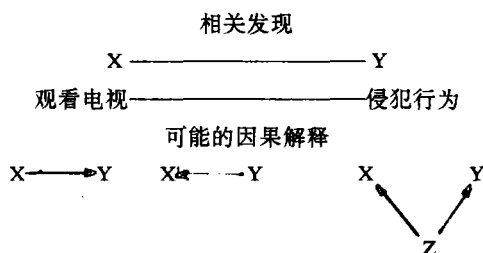
^① Rosenberg, M. J. 1965 When Dissonance Fails: on eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1:28-42.

^② Orne, M. T. 1962 On the Social Psychology of the Psychological Experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, 17:776-783.

^③ Shaver, K. G. 1981 *op cit* p38.

^④ *Ibid.* p39.

击行为)之间有关。然而这种结论只是一种相关解释,或者相关发现(correlational finding)。并且其中至少有二种可能的因果解释预含并不是电视节目导致儿童的攻击行为,见下图。



为了进一步确证具体的因果逻辑,社会心理学家就可以在实验室中实施严格的控制,以凸现自变量和因变量之间的精确的关系。例如 Lieber 等在实验室中把男孩或者女孩(被试)暴露在两种实验条件下,一种是观看摘录自暴力电视节目的电影片断,一种观看激烈的赛车比赛的片断,结果显现观看暴力电视的孩子更为频繁地按动一个特殊的红色按钮——它被假想为能对另一孩子实施痛苦(实质上,没有另一个孩子,也没有人实际上受到伤害)。因此这个实验就能够确证电视能成为儿童攻击行为的一个原因^①。

实验室情景是对日常生活情景的模拟,或者说,它是一个缩微的实在(miniature reality)^②,往往能在一定程度上表征真实的社会情景。但实验室实验在借助严格的实验控制而获得高度的研究内在效度外,其外在效度也在同等程度上丧失了,因为在实验室这种特殊的情景下,主试——被试之间的相互作用,被试的代表性,以及被试意识到正被研究而发生的成为好被试的角色扮演,严重干扰了实验结果的概括性。

四、对实验研究的反思

实验研究除却在社会心理研究的反应性之外,它还在几个方面受到严厉的审查。

1. 实验被试的代表性

正如 Rosenthal 等所指出的那样,人类的行为科学仍局限于以二年级大学生(sophomores)作为被试:他们一方面在一些心理学课程中注册,另一方面作为自愿者参与实验研究^③。很显然,大学生被试作为一个样组并不能指征整个的人类整体,或者说,他们并不具有代表性,因此,以之为对象的实验发现及结果的社会心理理论没有多少代表性。

另一位第一流的社会行为理论家 Gergen 从历史的社会心理学的角度深刻地洞悉了以大学生被试作为样组的社会行为论专注于共时分析,而抛开了丰富多彩的历史维度——即人类行为随着时代变化而变化的深刻内涵^④。当然,这些批评无疑有其深刻之处,但我们在批评一种事物时而不能设想另一种更为合理的选择来替代它,那么其意见的中肯性会失去诸多效用。大学生被试的代表性固然缺乏代表性,但从实验研究的费用,实验过程的管理等诸多方面,仍

^① Liebert, R. M. et al 1972 Some Immediate Effects of Televised Violence on Children' s Behavior. *Developmental Psychology*, 6, 469-475.

^② Myers, P. G. 1987 *Social Psychology, 2nd ed.* Mc Graw-Hill Book Company p16.

^③ Rosenthal, R., and Rosnow, R. L. 1969 The Volunteer Subject, pages 59-118 in Rosenthal, R. et al (eds), *Artifact in Behavior Research*, New York Academic Press. pp59-60.

^④ 参看 Gergen, K. J. et al (eds) 1984. *Historical Social Psychology*, Hillsdale, New Jersey, Lawrence.

有许多方便的地方。

2. 实验情景的人工性

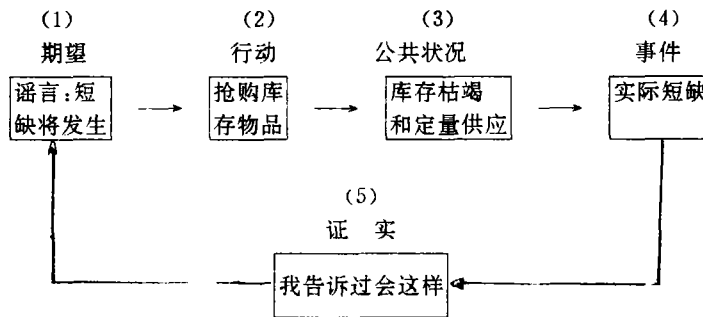
许多批评家注意到作为缩微实在的实验情景固然精致地模拟了自然的社会情景,但毕竟还是抛开活生生的社会互动的丰富内容。Harre 和 Secord 注意到实验情景破坏了社会互动的自然本质,使置于实验情景中的被试有种倾向作出符合该情景的社会行为表现,因此,以实验结果而作出的概括化是不精确的^①。因此他们偏好在自然情景下人种学方法论(ethnomethodology)研究。当然我们现在已经知道他们的批评不过在强调实验研究的外在效度的同时忽视了实验研究的独特性——它恰恰以高内在效度作为补偿。

另一些第一流的社会行为学家为实验室研究进行辩护。他们严格地在实验现实性(experimental realism)和生活现实性(mundane realism)之间进行区分^②。如果一个实验对被试有影响,迫使他认真对待并进入到实验程序中,那么就可以说,这个实验已具有了实验的现实性;而实验室情景与外部世界通常所发生的事件之间的相似程度问题,可称之为生活的现实性。因此,对实验情景并没有反映真实的现实情景的批评,在于把实验的现实性与生活的现实性相互混淆了。一个处于非现实性情景中的对象,并不会具有在 Milgram 的实验情景中的那种紧张不安的反应。但这种行为在 Milgram 的被试身上所发生,则是真实的——虽然日常生活中这些事情并不发生在他们身上。因此这种推论是可靠的:如果现实世界中真的发生类似情况,Milgram 的实验结果则可以精确而合理地说明人们对它将如何反应。

3. 主试——被试互动

实验程序中,主试——被试实际上处于特定的社会互动情景中,他们怀着一个共同的期望(expectancy)——即对可能的实验结果的期望而结合于一起。这种互动中的期望对实验结果产生特定的影响。

第一流的社会理论家 Merton 首先注意到一种特定类型的期望,即自我实现的预期(self-fulfilling prophecy SFP)。用 Merton 的话来说,“自我实现的预期,就是开始时有关情境的错误界定激发一种新的行为,这种行为使原初的错误界定成为现实”^③。在详细界定自我实现的预期效果时,Merton 描绘了战争期望如何导致一个国家以一种会实际激发战争的方式与另一国家交互作用。可以用市场短缺来表明自我实现的预期发生的各个过程。



资料来源: Hollander, E. P. 1981, p26.

Merton 所描绘的 SFP 只是种宏观的社会现象。它可以作为一种模式用来解释许多相关的

① Harre, R. and Secord, P. 1972 *The Explanation of Social Behavior*. Oxford: Basil Blackwell

② Aronson, E et al 1985 *Experimentation in Social Psychology*. In Ludzey, G and Aronson, E. (eds) *Handbook of Social Psychology*, Vol. 1. Hillsdale, N. J.: Erlbaum.

③ Merton, R. K. 1948 *The Self-fulfilling Prophecy*. *Antioch Review*, 8:193-210. p195. [Also in Merton, R. K. 1957] *Social Theory and Social Structure*. Glencoe, [11. Free Press. pp421-436].

政治及经济现象。社会心理学家在更为微观的层面上发现了主试——被试互动过程中的期望效应。

由 Rosenthal 所主持的一项著名的研究项目表明,主试对于研究结果的期望会无意,但明显地影响动物和人类的行为表现。Rosenthal 在一项迷津学习(maze-learning)的实验中告知充当实验助手的学生:他们要注意迷津学习中聪明的老鼠(maze-bright rats, MBR)和迷津学习中愚钝的老鼠(maze-dull rats, MDR)之间的行为差别,但实际上这些标签是随意乱贴的。结果表明 MBR 比 MDR 表现更好^①。很显然标签所承受的期望影响了学生的表现,也依次影响了老鼠的表现。尽管这个实验有一些不精确的实验条件的缺陷,但主试期待效应(experimenter expectancy effect)确实作为一种事实被发现,并且存在于几乎所有的实验室研究中。

随后, Rosenthal 等把同样的程序用于教学情景中。在研究开始时,研究者告知一个初级学校的教师们给其学生实施一项新的智力测验(IQ test),并且表明一些学生——“进步者”(bloomers)将会在一个学年中智力方面会有显著的提高。而事实上“进步者”这个标签是由研究者随机分配的。所有的学生在学年开始及结束都实施智力测验。结果显示,被标上“进步者”的学生比其他在 IQ 上有显著的提高^②。研究结果显示了教师对学生的期望效应或者称之为皮格马利翁效应(pygmalion effect, PE)。

重复 PE 的尝试产生一些不一致的结果。期望效应更容易在学业表现如数学和阅读测验中发现,而在智力测验中则较难发现。尽管如此,在 15 年的研究评论中, Rosenthal 等得出结论:在他们评论的 345 个研究中,有 2/3 发生了某种教师的期望效应^③。

到目前为止,我们并没有建构一种完美的社会心理研究法——它同时能良好地满足内在效度及外在效度。意识到这一点,就会告诫我们没有科学的方法,只有应用于科学中的方法。在评价一项研究结果或者一项研究程序时,我们应抱有更为审慎的态度。对方法多元性的宽容应该融入我们的理智之中。与其它方法相比,实验尤其是实验室实验有其独特的长处,这也使它占据独特的地位;同时在实验研究中有多种难以克服的误差可能介入。在意识到社会心理及行为研究的复杂性及艰苦性的同时,我们应该防止对于实验过份的乐观主义或者悲观主义。作为一种方法,实验无法取代它种方法,也不可能被它种方法所取代。作为社会心理研究方法体系中的一员,实验,连同其它种种方法,共同丰富和推动着社会心理学的研究事业。

主要参考文献:

- (1) Hollander, E. P. 1981 Principles and Methods of Social Psychology (4th ed.) N. Y. : Oxford.
- (2) Lindzey, G., and Aronson, E. (eds) 1985 Handbook of Social Psychology, Vol. 1, Hillsdale, N. J. : Erlbaum.
- (3) Milgram, S. 1975 Obedience to Authority. N. Y. : Harper and Row.
- (4) Rosenthal, R., and Foole, K. L. 1963 The Effect of Experiment Bias on the Performance of the Abino rat. *Behavioral Science*, 8:183-189.
- (5) Shaver, K. G. 1981 Principles of Social Psychology, (2nd ed). Cambridge, Mass. : Winthrop Publishers, Inc.

责任编辑:张宛丽

① Rosenthal, R., and Foole, K. L. 1963 The Effect of Experiment Bias on the Performance of the Albino rat. *Behavioral Science*, 8:183-189.

② Rosenthal, R., and Jacobsen, L. 1968 Pygmalion in the Classroom; Teacher Expectation and Pupils~Intellectual Development. New York: Holt, Rinehart and Winston.

③ Rosenthal, R., and Rubin D. B 1978 Interpersonal Expectancy Effects; The first 345 Studies. *Behavior and Brain science*, 3. 377-415.