

互联网社会科学实验：方法创新与价值评价

郝龙

(南京航空航天大学人文与社会科学学院, 江苏南京, 211100)

摘要: 实验法是社会科学研究和检验因果关系的最重要手段。随着在线行为与在线现象重要性的不断提升, 实验方法同互联网技术相结合逐渐成为研究方法创新的主要方向之一。互联网实验是一种在互联网空间中开展随机控制实验的研究方法。它遵循着实验方法的基本逻辑与研究流程, 同时在研究设计、实验环境、实验操作和受试者获取与分组等方面表现出一定的方法创新性。同传统社会科学实验方法相比, 互联网实验有着拓展研究对象、综合研究方式、扩大和丰富受试者的规模与类型、提升研究效率等优势, 但也存在着较高的技术门槛和测量与计算上的部分困难等局限。在复杂性科学和计算社会科学的方法体系中, 互联网实验还兼具实现“计算—模拟—实验”“微观—宏观”之间衔接与整合的重要方法价值。

关键词: 社会研究方法; 实验法; 互联网实验; 复杂系统; 计算社会科学

中图分类号: C910

文献标识码: A

开放科学(资源服务)标识码(OSID)

文章编号: 1672-3104(2020)06-0163-12



自实证主义方法论成形以来, 实验方法一直被视作社会科学探索和检验因果关系的最重要手段^[1]。虽然利用新开发出的计量技术(如倾向值匹配、工具变量), 社会科学家们已经有能力回答“自然发生的数据变化引起的因果问题”^[2], 但这些技术在控制无关变量干扰和确定变量时序方面依然远逊于实验法。当前, 根据实验环境与操作方式的不同, 常用的社会科学实验方法主要包括三种类型: 一是为保证内在效度而在纯粹人工创设环境中开展的实验室实验(Laboratory Experiments), 二是为保证外在效度而将自然事件当作干预的自然实验(Natural Experiments), 三是在真实社会环境中开展的, 致力于平衡内外效度的田野实验(Field Experiments)。随着互联网普及率和各种数字设备使用频率的提升, 日常生活的网络化与数字化趋势对上述实验方法均提出了不同程度的挑战。一方面, 网络世界的连通性、匿名性、脱域性等特征, 削弱了依赖“方便样本”

“人工模拟环境”的实验室实验的适用性; 另一方面, 网络世界作为现实社会结构的一部分, 也为田野实验创造了一种去身体化与数字化的全新田野环境。尤其是 Web2.0 技术出现以后, 互联网不再停留于扮演线下生活的信息记录空间角色, 它同时还搭建起一个新型的社会行动空间。发生于这一空间中的在线心理与行为现象, 在规则与逻辑、情境与表征等方面, 可能迥然不同于其线下形态。面对环境与研究对象的双重变化, 一些研究者开始尝试利用新兴的信息技术工具对传统实验方法进行改造、创新, 并由此发展出一种新的社会科学实验方法——互联网实验。

一、互联网实验方法概述

20 世纪 90 年代, 随着万维网(World Wide Web) 及其超文本标记语言(HTML)和 JavaScript

收稿日期: 2020-01-03; 修回日期: 2020-05-04

基金项目: 中央高校基本科研业务费专项“计算社会科学研究方法的创新、应用与价值研究”(ND2020006)

作者简介: 郝龙, 山东新泰人, 博士, 南京航空航天大学人文与社会科学学院公共管理系讲师, 主要研究方向: 网络社会学与计算社会科学, 联系邮箱: haolong2019@nuaa.edu.cn

脚本语言技术的出现,利用互联网开展社会科学实验的设想逐渐萌生出来。1995年,加拿大麦吉尔大学的韦尔奇(N. Welch)和美国汉诺威学院的克兰茨(J.H. Krantz)合作开展的心理学听觉实验,第一次将互联网与实验方法相结合^[3]。两年后,由克兰茨及其同事所做的“女性吸引力影响因素”研究则被视为第一项真正的互联网实验,他们将相同的实验方案在实验室和互联网上分别实施,而最终结果显示两种方法在研究结论上有着高度的相似性,由此证明了互联网实验作为一种研究方法的可行性^[4]。2000年,由伯恩鲍姆(M.H. Birnbaum)主编的《互联网上的心理学实验》(*Psychological Experiments on the Internet*)一书出版,书中简单回顾了互联网实验的初兴历程,并对数据有效性、方法优劣势、技术支持及其同实验室方法的异同等进行了介绍^[5]。此后,安德哈博(V. Anderhub)等将互联网实验引入个体经济决策行为研究之中^[6],由此推动这一方法由心理学向其他社会科学领域扩展。不过在这一时期,互联网实验尚未发展成为一种独立的研究方法,研究者们关注的焦点也主要放在通过互联网来丰富受试者的来源及其异质性方面,因此在名称上更多地使用“互联网实验”(Internet Experiment)^[7]、“万维网实验”([World Wild] Web Experiment)或“基于网络的实验”(Web-Based Experiment)^[8]。

进入21世纪后,互联网普及率和使用率的双重提升以及数据处理技术的快速发展,为样本代表性、实验平台、实验环境和数据处理等实验条件和手段工具的改善提供了支持,由此推动互联网实验方法进入一个新的发展阶段。2006年,萨尔加尼克(M.J. Salganik)等人的“社交因素对文化产品成功的影响”实验^[9]以及海德斯托姆(P. Hedström)的相应评论文章^[10],使互联网实验方法第一次出现在《科学》杂志之上。在萨氏等人的实验中,受试者全部从一个特定的在线音乐平台招募而来,而实验环境则以人工模拟的方式加以创建,向我们呈现了“虚拟实验室”的可行性和重要价值。此后的十年间,《Science》和《Nature》又接连刊发了多篇相关论文,引发了西方学界对这一新方法的关注。随着方法的日趋系统化,学术

界对互联网实验的称呼也开始发生变化,更多地使用“虚拟实验”(Virtual Experiment)^[11]和“在线(田野)实验”(Online [Field] Experiment)^[12]来指称。与Internet、Web等强调连通性的词汇相比,Online一词更能突出研究对象的行为属性和交互属性。

简单来说,互联网实验就是一种在数字化的互联网空间中开展随机控制实验,以检验变量间因果作用关系的研究方法。作为一种实证主义量化方法,它既遵循着“假设检验”的一般逻辑,也遵循着“刺激-反应”和“操纵-控制”等实验方法的特殊逻辑。受此规定,实验方案的“理论先行”、受试者的“随机分组”、实验条件的“高度控制”和刺激变量的“人为干预”等构成了互联网实验的基本要求^[13]。在研究流程上,互联网实验与传统实验方法大体一致,首先从特定的研究假设出发,设计出相应的实验方案;然后根据研究目标与设计,选择合适的互联网平台并编写相应的实验程序;其次从选定的互联网平台上获取和随机分组受试者,并以计算机程序化或人工方式开展实验控制和干预,收集实验数据;再次对所收集的数据进行预处理,并基于对前测和后测数据的比较计算刺激变量的平均干预效应及其统计显著性;最后对实验的内在效度和外在效度展开系统评估,以检验实验结果的可靠性。

二、互联网实验的方法创新

互联网实验尽管在研究逻辑和实验流程层面同传统实验方法基本一致,但实验环境和技术工具等方面的变化使其在操作应用层面呈现出一定的方法创新性。这些创新集中表现在实验类型、实验平台、实验程序、受试者的获取与分组四个方面。

(一) 多样化实验设计

互联网实验的最基本形式为“单次单因素实验”设计,即设置一个控制组和一个实验组(只有一个自变量),在较短时间内只在同一网络平台上进行一次实验。要注意的是,由于互联网实验既不需要将受试者聚在一处,也不需要严格统一受

试者的参与时间，单组实验的时间相比于传统实验方法可以被显著延长。如果担心因时间过长而导致实验受到某些不可预见因素的干扰，还可以设计为“重复性实验”，即在实验方案和平台不变的情况下，在不同的时段内进行多次实验。例如，森托拉(D. Centola)的“健康行为传染”实验设计时长为7周，为排除干扰，实验基于不同的受试者重复进行了5次，5次的结果基本一致，从而有力地证明了实验结果的内在效度^[14]。对于那些需要考虑时间变量的研究，则可以采用“追踪实验”设计，即考察同一批受试者在较长时间内的连续变化。例如，萨尔加尼克等人的单次实验就耗时长达69天^[15]。如果还需进一步考察环境变量的影响，则可以采用“同步实验”的设计，即将同一实验在不同的互联网平台上同时进行。例如，范德赖特(R.A. van de Rijt)等人的“成功动力学”实验，就同时在众筹网站、消费者评论网站、维基百科和请愿网站等四个网络平台上展开，而四个平台的实验结果基本一致，显著提升了实验结论的外在效度^[16]。

得益于数字化设计的低成本(尤其是低时间成本)和更大的受试者规模，互联网实验降低了进行多因素实验、组合实验和因果中介实验的难度。当实验中作为刺激因素加以操作的自变量数目超过一个时，即构成多因素实验。相较于只有一个自变量的单因素实验，多因素实验不仅能够考察单个自变量的主效应，还能计算多个自变量之间的交互效应。在互联网实验中，无论是直接设计多因素实验，还是将初始的单因素实验扩展为多因素实验都变得更为简单。例如，在一项关于“网络结构与合作贡献度”的实验中，研究者首先考察了网络集中程度对个体合作贡献度的影响，但发现这种影响并不显著。为检验这一结论，研究者在网络结构中额外加入了极端合作者(贡献度要么全额要么为零)这一新刺激变量，尽管结果仍不具有统计显著性，但这种设计方式的确有助于丰富和深化对变量间作用关系的认识^[17]。与只有一个因变量的多因素实验不同，组合实验设计是对多个考察不同因变量的实验方案的嵌套组合，即后一项实验的刺激变量内含在前一项实验的刺激变量之中，而前后两项实验分

别指向不同的结果变量。例如，霍波克(A. Coppock)等基于Twitter的“在线参与请愿”研究，就以在线传播的动员信息为共享刺激变量设计了两项实验。其中，第一项实验考察信息传播方式对受众参与行为的影响；第二项实验紧接着在此基础上，考察所传播的信息内容对受众信息转发行为的影响^[18]。这种组合实验设计的价值，在于以较低的时间和资金成本，通过一次实验完成对多个研究假设的检验。此外，组合实验设计还可以向因果中介实验转换，即通过对实验方案的组合嵌套发现多个变量间的因果作用链条。因果中介实验至少检验三个变量间的因果关系，其中先要检验自变量同中介变量和因变量的关系，然后再检验中介变量与因变量的关系。在具体操作上，因果中介实验可以采用双随机设计(两个实验方案的嵌套组合)、并发双随机设计(自变量取值数乘以中介变量取值数的多因素实验设计)和平行设计(前两种设计类型的综合)等不同形式^[19]。

互联网实验一般采用单盲实验设计，即只让实验的执行者了解实验设计，而向受试者隐匿实验的有关信息，其目的在于避免因受试者知晓自己在参加实验而出现“霍桑效应”(Hawthorne Effect)和“约翰亨利效应”(John Henry effect)等损害实验内在效度的现象。前者是指受试者发现自己正在被别人关注或者观察时，刻意改变自身的行为或态度；而后者则是指受试者在知晓分组状况时，可能会将实验视为“实验组”与“控制组”之间的一场竞赛，从而把主要精力放在比赛而非实验上。在实际操作中，互联网实验设计上的“单盲”，往往还能达到“双盲”的效果。因为很多互联网实验是通过计算机程序自动完成与受试者互动的任务，所以即使在“单盲”的设计下也能够有效避免由研究者与受试者人际互动所引发的“期望效应”(Expectation Effect)——研究者对实验结果期望的直接表达或间接流露，会诱导受试者做出迎合这一期望的回应)——以及干预分配的个体间不均衡等问题。

(二) 数字化实验平台

在互联网实验方法兴起之初，研究者主要通过自建在线平台的方式来进行实验。自建平台的

优势在于研究者能根据实验方案来布局平台结构和设计程序代码,既省去了获取平台使用权限的麻烦,也增强了对实验条件的控制力。然而,这种做法的劣势也非常明显,即需要耗费较多的人力与经费,且难以保证受试者的规模及其代表性。因此,新近的研究更倾向于使用那些用户规模较大的既有互联网平台。

目前,适用于互联网实验的在线平台大体可分为三类:第一类是在线众包平台,如美国亚马逊公司的机械顽童(Mechanical Turk)和中国的威客网。“众包”(Crowd sourcing)是一种基于互联网的连通性优势而形成的全新任务处理模式,它将那些原本为特定机构专业人员所处理的任务,通过互联网有偿外包给机构之外的众多非专业人员来集体协作完成,从而能够以极低的成本获取与调用机构之外的丰富资源(尤其是智力资源)^[20]。该模式所具有的低成本、大规模和高回报优势,使其非常适用于开展互联网实验。研究者可以将实验内容以有偿任务的形式发布在众包平台上,而承包者对任务的处理也就构成了对实验的参与^[21]。第二类是以脸书(Facebook)、推特(Twitter)、新浪微博为代表的在线社交平台^[22]。这类平台有着三大独特优势,即用户规模、社交关系网络、人均活跃度。知名的在线社交平台都具有跨地域、跨年龄甚至跨文化分布的庞大用户群,这不但可以提升样本数据的代表性,更使开展大尺度互联网实验成为可能。真实且大规模的在线社交关系网络,特别有助于开展诸如社会传染之类的基于人际关系的实验研究^[23]。较高的人均活跃度,使社交平台持续不断地生产与记录着丰富的个体心理与行为信息,为实验结果计算提供了必要或额外的数据支持^[24]。第三类是专业信息服务平台,如健身网站、婚恋网站、音乐网站、众筹网站、消费者评论网站、维基百科和政务公开网站。这些专门化平台聚焦于特定的专业/兴趣领域,特别适用于那些需要特殊情境或特定人群的互联网实验。这类平台有时也具备社交功能,虽然在关系网络的规模方面远逊于在线社交平台,但其所建立或维持的社会关系通常带有一定的专业性和更明确的目的性。

在具体的互联网实验中,研究者对在线平台

的选择需要遵循契合性和可用性两大原则。所谓契合性原则,是指在线平台的结构与特征应当符合实验方案的需要。一方面,不同类型的平台不仅会塑造出差异化的实验环境,也限定着实验可以利用的资源与条件。以在线动员实验为例,那些以请愿、联署、声援和投票等低成本参与行为作为结果变量的实验,可以在开放的大型社交网络平台上进行;但当实验中涉及金钱(如在线筹款实验)之类高度需要机构信用做背书的因素时,专业的信息服务平台(如非营利组织的官方网站)相比之下则更为合适。另一方面,以往研究也发现,用户特征与在线平台特征之间存在着不同程度(或强或弱)的相关关系,这意味着不同类型的实验平台在受试者库的规模、结构与特征等方面常常存在着较大的差异,需根据实验方案所要求的受试者群体类型进行选择。至于可用性原则,则是强调研究者必须能够获得在线平台的使用权限。相较而言,自建平台和众包平台最容易使用,而各类社交平台和专业平台由于为商业公司和社会组织机构所拥有,获取使用权限的难度较大。

(三) 程序化实验操作

出于理解上的便利性,初始的互联网实验方案通常是以自然语言的形式表述出来。然而,自然语言无法被计算机直接识读与执行,必须转换为二进制代码形式的程序语言,即通过编写一系列的功能程序,将那些原本需要人工操作的实验任务交由计算机完成。干预、控制与测量是互联网实验程序必须具备的三项核心功能。实验干预程序主要执行向受试者施加实验刺激的任务,其是否需要单独编写要视在线平台的可用条件而定。考虑到实验环境的“自然发生”要求与经济成本,最理想的情况是直接以在线平台的既有功能程序来实施干预,其次是根据实验方案对既有平台程序做一些修改后加以利用。例如,有研究者就曾以社交平台中常见的“匿名浏览”(查看他人页面而不留下任何痕迹)功能作为刺激变量,以考察匿名功能对“弱信号”能力的剥夺是否会导致约会成功率的下降^[25]。实验控制程序用于对实验发生的环境条件进行控制。该程序一方面要保证全体受试者处于大致相似的在线环境中,

并尽量将那些有可能对实验结果产生干扰的因素排除出去；另一方面，也要避免控制组直接或间接受到刺激变量的影响。无论在何种在线平台上，这部分程序都需要单独编写，并嵌入平台的系统架构之中。实验测量程序的功能在于收集实验过程中所产生的各种数据，尤其是结果变量的指标数据，并将其存入特定的数据库中。这一程序通常也需要根据在线平台的API接口命令而单独编写，因为如果直接使用平台本身的数据存储程序，很可能导致大量无关数据混入实验数据库中，给后期的数据处理增添不必要的麻烦。

目前已有多款适用于主流在线平台的程序设计工具可供选用，为社会科学研究者编写实验程序提供了便利。例如，针对亚马逊的机械顽童平台，就有支持实验迭代计算的Turkit工具^[26]，支持同步实验和追踪实验计算的TurkServer工具^[27]和用于经济学博弈行为研究的Seaweed工具^[28]。其中，由MIT开发的Seaweed程序使用可视化界面进行实验设计，非常适合那些缺乏编程经验的研究者。著名社交网站Facebook也提供了一款名为Planout的在线实验设计工具，该工具能够将实验设计代码与应用程序代码分离开来，使研究者只需通过编写一系列简单的指令脚本，即可实现对页面参数的设置与调节^[29]。此外，Google Forms、LimeSurvey、Qualtrics和问卷星等工具，能为那些需要使用问卷和量表测量的互联网实验提供必要的服务支持。当实验程序设计完毕并嵌入实验平台之后，研究者还需对程序进行调试和完善，着重检查以下三个方面的情况：一是实验干预的作用效果，防止刺激变量无效(未能进行干预)或受损(未能达到预期的完整干预作用)等问题的出现；二是实验条件的控制效果，防止干扰变量的出现和刺激变量的溢出(实验干预不仅作用于实验组，还以直接或间接方式作用于控制组)；三是实验结果数据的记录，防止出现信息记录错位、数据存储中断、数据格式错误等问题。

(四) 获取与分组受试者

传统实验方法主要通过“招募”的方式来获取受试者，这一方式对互联网实验而言同样适用。最理想的招募方法是以预先获得的受试者库名单作为抽样框，采用概率抽样方法从中随机抽

取出一定数量的受试者。当受试者库规模足够大，且其分布结构与总体人口结构相类似时，这种方法能够同时保证样本的代表性及其获取的随机性。然而，受成本等因素限制，以上条件通常难以获得，很多实验只好转而采用非概率的便利抽样来获取受试者。便利抽样的优势在于节约成本、方便快捷，但同时也有着巨大的劣势，即由于无法保证样本获取的随机性，实验数据中可能会掺杂进样本的自我选择偏误。除此之外，抽样的对象也一般局限于大学生群体，对部分研究主题而言，这类样本的代表性被认为会有损实验的外在效度^[30]。

在上述两种方法之外，互联网实验利用互联网的连通性优势，创造出了第三种招募方法——众包招募。所谓众包招募，是指研究者将实验任务放置在专门的众包网站上，通过提供一定的报酬引导受试者为获得酬金而主动去完成这些任务^[31]。研究显示，使用众包方法招募受试者并不会损害样本的代表性^[32]。随着网络普及率的提升，网民群体结构与总体人口结构之间的相似性越来越强；即使单一网络平台用户的分布结构难以与总体人口结构相符合，但其多样性和代表性仍要优于以往常用的学生样本或其他便利样本^[33]。使用招募方式来获取受试者，除了需要注意样本选择偏误问题之外，一般还要考虑向参与者提供一定的报酬。由于互联网的去身体化特性，研究者对受试者的控制力被严重削弱。一旦缺少足够的激励，受试者很容易在实验过程中大量流失^[34]。因此，以众包方式招募受试者，其规模势必会受到实验经费的制约。

为了应对选择性偏误与规模限制问题，互联网实验又创造性地发展出一种被称为“在线抓取”的新方法，即研究者在用户不知情的情况下，利用平台管理员权限直接从中抓取受试者及其相关数据信息。与招募方式相比，这种方法不仅简单便捷，而且更容易获取规模庞大的受试者群体。在一项关于选举投票的在线实验中，研究者就曾通过抓取方法获得了六千多万的受试者^[35]。受试者规模的增大，一方面能够避免“招募”方式带来的样本自我选择偏误，并为受试者的随机分组提供必要前提条件；另一方面，也能够显著

提升样本的代表性,优化实验的外在效度。然而,研究伦理(知情同意)与法律(个人隐私权利)的双重责难,以及商业平台管理员权限的获取困难,都会降低抓取方法的易用性。

在获取足够数量的受试者后,紧接着就要对其进行分组。由于无法同时测量到同一位受试者在接受干预与未受干预两种条件下的变化,社会科学实验总是在反事实框架(Counterfactual Framework)下进行,实验结果需要通过计算刺激变量的平均干预效应来获得,因此保证实验组与控制组之间的可比性(相似性)至关重要。由于当样本数量较少时不具备随机化的必要条件,传统实验方法通常采用“配对法”来分组:根据预先确定的少数几个重要特征,将特征相似的一对受试者分别划入实验组和控制组。配对法实际上只能算是一种折中方法,因为“我们无从知晓哪些相关变量应作为配对的依据”^[36];而且受试者的复杂特征通常难被研究者完整把握,那些未被重视或识别的特征因素很可能对实验造成干扰。互联网实验对受试者规模的扩展,为随机分组方法的使用提供了必要的前提条件。不过新的问题是,受试者参与实验时间的不统一,仍有可能损害随机化的效果。当实验过程时长较短时,研究者在获取到所需数量的受试者后就立即进行分组,并不会因受试者参与时间的差异而造成组间特征分布的不均衡。但当实验时间较长(如30天以上)时,受试者可以在一开始就加入进来,也可以等到最后一天才参与。过长的时间间隔,很可能使后参与的受试者受到未经控制的因素(如其间恰逢某一重大社会事件发生)的影响。这种情况下,预先分组的方法就难以保证组间特征分布的随机性。一个可行的策略是,将“预先分组”变为“即时分组”,即事先不对受试者进行分组,只有当参与行为发生时才通过计算机程序将其随机分入实验组或控制组^[37]。

三、互联网实验方法的优势与局限

作为一种新的网络化和数字化实验方法,互联网实验在研究对象、实验环境与操作技术等方

面的诸多创新,使其在扩展研究对象、综合研究方式、扩大和丰富受试者的规模与类型等方面,相对于传统实验方法而有着一定的方法优势(见表1)。

首先,互联网实验能够丰富和拓展实验法的研究对象和适用范围。随着网络交互技术的创新发展和广泛应用,互联网数字化空间的双重属性愈加鲜明:一方面,它是一个信息记录空间,是线下物理空间的复制或延伸;另一方面,它又是一个新型的社会行动空间,在线活动的频率、范围和重要性日益提升。得益于这种双重属性,互联网实验方法既能够应用于线下心理与行为的研究——在这方面它是传统实验方法的补充,也可以用来研究新兴的在线心理与行为以及线上线下的相互作用关系——在这方面它有着不可替代的重要性,更加凸显出其独特的方法价值。

其次,互联网实验能够实现以往多种实验类型的有效综合。实验环境的可控性和仿真性,分别关系到研究结果的内在效度和外在效度。但在实际操作中,上述两项标准呈现一种矛盾对立状态,对其中一项的强化会造成对另一项的弱化。在互联网实验中,这一问题有机会得到一定程度的解决。首先,我们能够通过创建虚拟实验室的方式来确保环境的可控性。虚拟实验室可以被视作传统实验室的数字化呈现,其实验环境仍通过人工模拟方式创设,对干扰的控制力较强,适用于讨论一般性议题。其次,我们也能够运用在线田野实验的方法来提升环境的仿真性。在线田野实验将实验放置在自然发生的互联网环境之中,对干扰的控制力相对减弱,但随着真实情境化因素的加入,其外在效度会得到显著提升。最后,数字化设计的低成本和受试者获取的便利性,还降低了实验方案平行设计的难度。所谓平行设计,是指可以同时设计基于人工环境和田野环境的两套实验方案,最终通过对研究结果的比照来平衡实验的内外效度。

再次,互联网实验能够有效扩大受试者的规模,并提升其异质性和代表性。借助互联网的优势,研究者能够通过多样化的在线招募方法与程序抓取方法,以较低的时间与经费成本,获得规模更大、稳定性相对更高、更具多样性且跨文化

分布的受试者群体。受试者规模的扩大，有助于提升样本数据的代表性，进而优化实验结果的外在效度。除此之外，互联网还为接近特定隐秘性社会群体和亚文化群体提供了新的途径，也为跨文化研究提供了便利。在以往的社会科学实验中，由于文化因素过于宏大和抽象，难以被有效操作。互联网实验凭借受试者的跨地域、跨文化分布，能够将这一软变量有效引入实验研究之中。

最后，程序化设计能够显著提升实验研究的效率。互联网的数字化特征，使实验过程可以通过计算机程序自动运行，在节省人力的同时，也有助于减少乃至避免研究者对受试者可能造成的干扰；实验的测量也可以不再需要人工完成，流处理技术能够将实验过程中产生的数据实时记录下来。除了节省实验室空间、人工工时、设备和管理费用之外，互联网实验还能够实现理论与实验间的快速迭代。在以往的实验研究中，从提出理论假设到得出实验结论再到修正理论假

设，以及再次对新假设进行验证，往往需要经历一个较长的时间过程。互联网实验的快速和便捷，有效缩短了实验与理论间的迭代过程，这无疑会大大提高研究的效率。

当然，互联网实验也存在着一定的方法局限性。一方面，实验环境和实验手段/工具的数字化和程序化，为互联网实验设定了较高的技术门槛。在应用这一方法时，如何将自然语言表述的实验方案以计算机语言的方式加以实现，如何确保实验程序能够完整、稳定地实现预期的干预和控制效果等都是需要妥善解决的技术问题，都向缺乏计算思维和编程经验的社会科学研究者提出了较高的跨学科合作的要求。此外，绝大多数的互联网服务平台均由商业公司运营和掌控，其后台管理权限并不对外开放，这就为那些以在线田野实验方式设计实验方案的研究者提出了同商业公司沟通和合作的要求。另一方面，互联网实验在数据测量方面也存在着一定的困难。“前测”加“后测”是现代实验方法的基本搭配，实

表1 四种社会科学实验方法比较

实验操作	实验室实验	田野实验	自然实验	互联网实验
研究对象	一般化线下心理与行为现象	情境化线下心理与行为现象	情境化线下心理与行为现象	一般化或情境化的线上或线下心理与行为现象
实验环境	人工模拟的抽象社会环境	自然发生的具象环境	自然发生的具象环境	自然或人工的互联网环境
受试者库	大学生	各类人群	各类人群	特定互联网平台的用户
受试者招募方式	校内招募	校外招募	无需招募	在线招募或抓取
受试者规模	数人到数十人规模	数十人到数百人规模	数十人到数百人规模	规模极大，从千人到千万人
受试者分组	预先分组	预先分组	自然分组	预先分组或即时分组
受试者是否知晓参与实验	知晓	可知晓也可不知晓	不知晓	可知晓也可不知晓
实验控制	高度控制	控制力弱于实验室实验，但强于自然实验	不施加实验控制，分析时用计量方法控制	控制力弱于实验室实验，但可通过增大受试者规模来降低变异性
实验干预	人为干预	人为干预	自然发生	人为干预
实验流程安排	人工操作	人工操作	无	机器程序操作
实验时长	较短，一般以小时计算	时间较长，一般以天计算	事件发生之后	时间可短可长，可按天或月计算
实验测量	前测与后测	前测与后测	后测	前测与后测，过程数据的实时记录

验结果的计算也是通过对前后测数据的比较来量化呈现刺激变量的平均干预效应。然而,在一部分互联网实验中,前测任务可能难以完成,例如当结果变量被视为纯粹的计算机程序响应结果时,在受试者同程序交互之前,研究者无法预先获取相应数据;当受试者群体规模过于庞大时,通常也难以进行前测。与此同时,许多互联网数据都以非结构化数据的形式存在,这给统计分析的信度和效度均带来了考验。当前,对音频和视频数据的常用处理方法是将其转换为文本数据;而对文本数据的处理方法主要有两种,一种是计算词频、词序、词距的计量方法,另一种则是基于机器学习的自动内容分析法。这两种方法在数据处理的精细度和准确性上均远逊于结构化数据处理方法,而且数据信息本身与变量操作化指标之间的匹配程度——数据效度问题——常常也存在着很大疑问。

四、复杂性研究中互联网实验的方法价值

互联网实验方法的价值,实际上不只表现在对于传统实验方法不足的改进方面。在数字化生活愈发重要和计算社会科学方法体系日臻成熟的背景下,互联网实验还有着实现“计算—模拟—实验”“微观—宏观”之间衔接与整合的重要方法价值。

(一) 计算社会科学研究方法之间的相互补充

兴起于20世纪80年代的复杂性科学(complexity science),极大地开阔了社会科学的视野,也使研究者深刻认识到:人类社会是一个具有自组织能力的复杂适应系统(complex adaptive systems),兼具简单性与复杂性、规律性与随机性、有序性与无序性,其演变过程有简单系统所不具备的模式和特征^[38]。面对这样的研究对象,尤其是面对适应性主体(adaptive agent)与环境以及其他主体的复杂、非线性交互作用关系,传统的实证研究方法遭遇了严重的适用性危机。这使一直深谙现有方法局限的社会科学家们更加迫

切地探索在方法体系上取得突破的路径。正是在这一背景下,推动社会科学同信息科学和计算机科学的相互融合,已成为量化社会研究方法创新的主要方向之一,并由此催生出一个名为“计算社会科学”的新兴交叉学科领域^[39]。计算社会科学通过将多样化的信息数据和数字化工具引入到社会研究之中,创造性地发展出社会科学大数据计算、社会仿真模拟和互联网社会科学实验三种研究方式,为提升量化研究的资料收集能力、认识能力和分析能力开辟出了新的方法路径^[40]。

计算社会科学的出现,最初源于社会模拟(social simulation)——当前最具代表性的是基于行动者的建模与仿真(agent-based modeling and simulation, ABMS)——方法的发展。该方法通过数学建模并转换为可被计算机执行的程序,尝试建立起一个与现实社会相平行的“人工社会”,借以考察现实社会中的复杂互动机制,预测社会变迁趋势等^[41]。不过,社会建模的困难和程序迭代的复杂,使得“人工社会”与“现实社会”之间始终存在着难以弥合的鸿沟,进而制约着模拟方法在社会科学研究中的应用^[42]。此后,随着互联网、移动互联网和物联网普及率与使用率的相继提升,海量的人类心理与行为互动信息源源不断地在三大网络中生成与累积,为社会科学研究提供了全新的数据来源^[43];而大规模数据处理技术的进步也为数据密集型知识发现创造了可能,并由此为计算社会科学增添了新的内容——社会科学大数据计算^[44]。

社会科学大数据计算方法的出现,的确为在线心理和行为研究提供了新的数据来源和手段工具。然而,现实中的多数社会科学大数据普遍存在着价值低、可关联性差(信息维度少)、信息真实性存疑等潜在问题,加之社会系统本身的复杂性,大数据计算检验因果关系的能力也严重落后于其发现相关性知识的能力^[45]。一部分学者就此提出,在大数据时代探求“是什么”(相关)要比“为什么”(因果)更为重要。另一部分学者则持相反观点,认为因果关系依然重要,并试图去弥补大数据计算方法在这一方面的不足。前一种观点实际上并不足取,因为对社会科学研究而

言，尽管相关性知识对于加深对社会的认识意义非凡，但因果性知识才构成了社会理论建立、发展与检验的核心内容。然而后一种的观点亦不完整，因为实验方法才是社会科学进行因果关系检验的最常用和最重要手段。正是在这层意义上，互联网实验方法构成了对大数据计算和社会模拟方法的补充^[46]。例如，大数据计算方法的一大特征是数据生产与科学研究之间的分离，即数据不再是通过研究工具测量产生的结果，而是在研究之外独立生成。这种数据生产的独立性，使得大数据计算普遍面临着数据信度和效度问题的威胁。互联网实验方法的出现，有助于重建数据生产与科学研究之间的关联，并通过重新引导数据的生成过程来检验大数据计算的结论^[47]。再如，在社会仿真模拟方法中，对社会行为与现象的建模精度直接影响着模拟的仿真程度。互联网实验方法所发现或检验的变量间因果关系，对于建模精度的提升而言大有帮助，尤其是在基于行动者的模拟中能够为交互行为规制的设置提供重要的参考^[48]。

（二）微观与宏观之间的方法连接

微观行为与宏观现象之间的关系，始终是困扰社会科学家的重要议题之一。按照复杂自适应系统理论的观点，要真正理解集体的属性，就必须考虑个体的行为及其相互作用；就必须考虑它们是如何聚合为集体属性的，亦即从微观到宏观的“涌现”过程。同时，集体属性对个人的行为也有着形塑、规制作用，因此需要了解集体属性在某个时间点上如何影响个人偏好、信念等，并如何进而影响个人的行为方式。这种影响可能反过来继续对集体属性产生某种作用，从而形成动态循环的复杂过程。传统的社会实证研究主要基于随机调查数据，这些数据虽然在探知个体的态度与行为方式等方面表现出色，但在分析和理解个体之间交互式作用过程方面则常常无能为力。在对这一问题的回应上，计算社会科学的三种研究方式均有着较大的方法价值。例如，基于行动者的建模与仿真方法，就是通过计算机仿真技术模拟行动者之间的持续互动，以获得对复杂社会现象及其演变过程的更好理解，探究从微观行为

到宏观现象的涌现过程。大数据时代到来以后，计算社会科学又尝试利用大规模的个体数据来分析和认识宏观社会现象。

长期以来，社会科学实验法都被认为是一种个体主义取向的研究方法，其分析单位局限于微观个体层面，所能获得的有关“总体”的知识也更多地表现为一种忽略单个个体之间差异的“平均”认识。这正是实验法尽管为心理学、经济学所重视，却始终为集体主义取向的社会学家们所忽视的原因所在。在这一方面，互联网实验方法庞大的受试者规模、多样化数据来源和自动化记录工具以及社会网络和交互因素的加入，为从微观到中观(甚至宏观)的因果知识发现提供了可能。例如，随着社会网络理论的发展，行动者之间的互动关系及其影响日益受到社会科学的关注。然而，由于受试者数量有限，加之关系变量常常难以控制和测量，传统实验方法在考察大规模受试者间相互作用方面普遍面临着障碍。互联网实验的出现，被认为有机会打破以往实验法只适于研究小团体互动的限制。基于庞大的受试者规模、低成本且便捷的即时在线交互和(大规模的)真实或模拟社交网络数据，互联网实验方法在研究在线社会互动、社会传染方面已取得了重要进展。正如海德斯托姆在评价萨尔加尼克等人在《实验宏观社会学》一文中所指出的那样，互联网实验方法向我们展示了如何利用互联网技术来克服传统实验的规模和信息限制，对理解从个人行为 and 相互作用产生集体属性结果的复杂过程具有重要的方法论意义^[49]。

参考文献：

- [1] FALK A, HECKMAN J J. Lab experiments are a major source of knowledge in the social sciences[J]. *Science*, 2009, 326(5952): 535-538.
- [2] LIST J A . An introduction to field experiments in economics[J]. *Journal of Economic Behavior & Organization*, 2009, 70(3): 439-442.
- [3] WELCH N, KRANTZ J H. The world-wide web as a medium for psychoacoustical demonstrations and experiments: Experience and results[J]. *Behavior Research Methods Instruments & Computers*, 1996, 28(2):

- 192-196.
- [4] KRANTZ J H, BALLARD J, SCHER J. Comparing the results of laboratory and world-wide web samples on the determinants of female attractiveness[J]. *Behavior Research Methods Instruments & Computers*, 1997, 29(2): 264-269.
- [5] BIRNBAUM M H. Psychological experiments on the internet[M]. San Diego: Academic Press, 2000.
- [6] ANDERHUB V, MÜLLER R, SCHMIDT C. Design and evaluation of an economic experiment via the internet[J]. *Journal of Economic Behavior & Organization*, 2001, 46(2): 227-247.
- [7] REIPS U D. Standards for internet-based experimenting[J]. *Experimental Psychology*, 2002, 49(4): 243-256.
- [8] MORONEY N, ESCHBACH R, MARCU G G. Unconstrained web-based color naming experiment[J]. *Proceedings of SPIE - The International Society for Optical Engineering*, 2003, 5008: 36-46.
- [9] SALGANIK M J, DODDS P S, WATTS D J. Experimental study of inequality and unpredictability in an artificial cultural market[J]. *Science*, 2006, 311(5762): 854-856.
- [10] HEDSTRÖM P. Experimental macro sociology: Predicting the next best seller[J]. *Science*, 2006, 311(5762): 786-787.
- [11] CHESNEY T, CHUAH S H, HOFFMANN R. Virtual world experimentation: An exploratory study[J]. *Journal of Economic Behavior & Organization*, 2009, 72(1): 618-635.
- [12] CENTOLA D. The spread of behavior in an online social network experiment[J]. *Science*, 2010, 329(5996): 1194-1197.
- [13] 郝龙. 互联网社会科学实验:数字时代行为与社会研究的新方法[J]. *吉首大学学报(社会科学版)*, 2018(2): 26-34.
HAO Long. Online social science experiment: A new research method of behavior and society in the digital age[J]. *Journal of Jishou University (Social Sciences)*, 2018(2): 26-34.
- [14] CENTOLA D. An experimental study of homophily in the adoption of health behavior[J]. *Science*, 2011, 334(6060): 1269-1272.
- [15] SALGANIK M J, DODDS P S, WATTS D J. Experimental study of inequality and unpredictability in an artificial cultural market[J]. *Science*, 2006, 311(5762): 854-856.
- [16] VAN DE RIJT R A, KANG S M, RESTIVO M, et al. Field experiments of success-breeds-success dynamics [C]// *Proceedings of the National Academy of Sciences of the United States of America*, Washington: National Academy of Science, 2014, 111(19): 6934-6939.
- [17] SURI S, WATTS D J. Cooperation and contagion in web-based, networked public goods experiments[J]. *Sigecom Exchanges*, 2011, 10(2): e16836.
- [18] COPPOCK A, GUESS A, TERNOVSKI J. When treatments are tweets: A network mobilization experiment over twitter[J]. *Political Behavior*, 2016, 38(1): 105-128.
- [19] 胡安宁. 社会科学因果推断的理论基础[M]. 北京: 社会科学文献出版社, 2015: 107.
HU Anning. The theoretical bases of causal inference in social science[M]. Beijing: Social Sciences Academic Press(CHINA), 2015: 127.
- [20] HOWE J. The rise of crowdsourcing[J]. *Wired Magazine*, 2006, 35(4): 996.
- [21] HORTON J J, RAND D G, ZECKHAUSER R J. The online laboratory: Conducting experiments in a real labor market[J]. *Experimental Economics*, 2010, 14(3): 399-425.
- [22] MUNGER K. Tweetment effects on the tweeted: Experimentally reducing racist harassment[J]. *Political Behavior*, 2017, 39(3): 629-649.
- [23] TSVETKOVA M, MACY M W. The social contagion of generosity[J]. *PloS One*, 2014, 9(2): e87275.
- [24] GOLDERS A, MACYM W. Digital footprints: Opportunities and challenges for online social research[J]. *Annual Review of Sociology*, 2014, 40(1): 129-152.
- [25] BAPNA R, RAMAPRASAD J, SHMUELI G, et al. One-way mirrors in online dating: a randomized field experiment[J]. *Management Science*, 2016, 62(11): 3100-3122.
- [26] LITTLE G, CHILTON LB, GOLDMAN M, et al. TurkIt: Human computation algorithms on mechanical turk[C]// *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*. New York, USA, 2010.
- [27] MAO A, CHEN Y, GAJOS K Z, et al. TurkServer: Enabling synchronous and longitudinal online experiments [C]// *In AAAI Workshop-Technical Report*, Vol. WS-12-08: 33-39.

- [28] CHILTON L B, SIMS C T, GOLDMAN M, et al. Seaweed: A web application for designing economic games [C]// BENNETT P, CHANDRASEKAR R. HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation, New York: Association for Computing Machinery, 2009: 34–35.
- [29] BAKSHY E, ECKLES D, BERNSTEIN M S. Designing and deploying online field experiments [C]// Chin-Wan Chung. WWW '14: Proceedings of the 23rd international conference on World wide web, New York: Association for Computing Machinery, 2014: 283–292.
- [30] GÄCHTER S. (Dis) advantages of student subjects: What is your research question?[J]. Behavioral & Brain Sciences, 2010, 33(2–3): 92–93.
- [31] SHANK D B. Using crowd sourcing websites for sociological research: The case of amazon mechanical Turk[J]. American Sociologist, 2015, 47(1): 47–55.
- [32] BERINSKY A J, HUBER G A, LENZ G S, ET A L. Evaluating online labor markets for experimental research: Amazon. com's mechanical Turk[J]. Political Analysis, 2012, 20(3): 351–368.
- [33] BUHRMESTER M, KWANG T, GOSLING S D. Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data?[J]. Perspectives on Psychological Science, 2011, 6(1): 3–5.
- [34] DUERSCH P, OECHSSLER J, SCHIPPER B C. Incentives for subjects in internet experiments[J]. Economics Letters, 2008, 105(1): 120–122.
- [35] BOND R M, FARISS C J, JONES J J, et al. A 61-million-person experiment in social influence and political mobilization[J]. Nature, 2012, 489(7415): 295–298.
- [36] 艾尔·巴比. 社会研究方法: 第十一版[M]. 邱泽奇, 译. 北京: 华夏出版社, 2009: 229.
- BABBLE E. The practice of social research[M]. Trans. QIU Zeqi. Beijing: Huaxia Publishing House, 2009: 229.
- [37] ARECHAR A A, GÄCHTER S, MOLLEMAN L. Conducting interactive experiments online[J]. Experimental Economics, 2018, 21(1): 99–131.
- [38] 米勒, 佩奇. 复杂适应系统: 社会生活计算模型导论[M]. 隆云滔, 译. 上海: 上海人民出版社, 2012: 32–36.
- MILLER J H, PAGE S E. Complex adaptive systems: An Introduction to computational models of social life[M]. Trans. LONG Yuntao. Shanghai: Shanghai People's Publishing House, 2012: 32–36.
- [39] LAZER D, PENTLAND A, ADAMIC L, et al. Computational social science [J]. Science, 2009, 323(5915): 721–723.
- [40] 乔菲-雷维利亚. 计算社会科学: 原则与应用[M]. 梁君英, 等译. 杭州: 浙江大学出版社, 2019: 1–2.
- CIOFFI R C. Introduce to computation social science: Principles and application[M]. Trans. LIANG Junying, et al. Hangzhou: Zhejiang University Press, 2019: 1–2.
- [41] GILBERT N, TROITZSCH K G. Simulation for the social scientist [M]. Maidenhead: Open University Press, 2005: 2–6.
- [42] BANKES S C, LEMPERT R J, POPPER S, et al. Making computational social science effective: Epistemology, methodology, and technology[J]. Social Science Computer Review, 2002, 20(4): 377–388.
- [43] GOLDBER S A, MACY M W. Digital footprints: Opportunities and challenges for online social research[J]. Review of Sociology, 2014, 40(1): 129–152.
- [44] 郝龙, 李凤翔. 社会科学大数据计算——大数据时代计算社会科学的核心议题[J]. 图书馆学研究, 2017(22): 20–29.
- HAO Long, LI Feng-xiang. Big data computing applied in social science: The core issue of computational social science in the big data age[J]. Research on Library Science, 2017(22): 20–29.
- [45] 郝龙. “计算”的边界: 互联网大数据与社会研究[J]. 中南大学学报(社会科学版), 2018(2): 20–29.
- HAO Long. The boundary of computation: Internet big data and social research[J]. Journal of Central South University(Social Sciences), 2018(2): 20–29.
- [46] 萨尔加尼克. 计算社会学: 数据时代的社会研究[M]. 赵红梅, 赵婷, 译. 北京: 中信出版社, 2019: 133–135.
- SALGANIK M J. Bit by bit: Social research in the digital age[M]. Trans. ZHAO Hongmei, ZHAO Ting. Beijing: China Citic Press, 2019: 133–135.
- [47] CHENG J, DANESCU-NICULESCU-MIZILC, LESKOVEC J, et al. Anyone can become a troll[J]. American Scientist, 2017, 105(3): 152–155.
- [48] 罗俊. 计算·模拟·实验: 计算社会科学的三大研究方法[J]. 学术论坛, 2020(1): 35–49.
- LUO Jun. Computation·Simulation·Experiment: Three research methods of computational social science[J]. Academic Forum, 2020(1): 35–49.
- [49] HEDSTRÖM P. Experimental macro sociology: Predicting the next best seller[J]. Science, 2006, 311(5762): 786–787.

Online experiment for social sciences: Methodological innovation and evaluation

HAO Long

(School of Humanities and Social Sciences,
Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China)

Abstract: Experimental method is the most important approach for social sciences to explore and test causality. With the constantly increasing importance of online behaviors and virtual phenomena, the combination of experimental method and Internet technology has gradually become one of the main directions of research method innovation. Online experiment, a research method to carry out random-control experiments in the digital space, follows the basic logic and research procedure of experimental methods and has certain methodological innovations in research design, experimental environment, experimental operation, acquisition and grouping of subjects. Compared with traditional experimental methods of social sciences, online experiment boasts of such advantages as expanding research objects, integrating comprehensive research methods, enlarging and enriching the scales and types of subjects, and improving research efficiency. However, there are some methodological limitations, such as higher technical threshold and some difficulties in measurement and calculation. In the method system of complexity science and computational social sciences, online experiment is also of important methodological values in realizing the integration of both "computation-simulation-experiment" and "micro-macro".

Key Words: social research methods; experimental method; online experiment; complex system; computational social science

[编辑: 游玉佩]