

# 生态学实验“可重复”困难的原因及对策

肖显静

华南师范大学公共管理学院, 广州 510006

**摘要** 生态学实验“可重复”存在诸多困难。在本体论上,主要有自然的变异性以及大尺度的限制等原因,对此,采取的对策是:或者使用易于处理的生物或生态系统来阐明相关过程,或者选择那些同质性的或平衡的系统进行研究,或者模拟自然进行微宇宙实验。不过,所有这些对自然的简单化方面的选择或者对自然的简化模拟,目的是获得“可重复性”高的实验,虽然一时有效,但并非总是可行。在认识论上,生态学实验对象的复杂性、有机整体性、历史性决定了对它们的相关认识的正确性受到限制,这直接影响到实验的“可重复”,为此,准确确定实验场所、清楚界定相关概念等就成为必需,由此能够达到生态学实验的正确性(实在性)与“可重复性”的双赢。在方法论上,不完整的实验报告以及缺乏相关的方法细节,是造成生态学实验“可重复”困难的重要原因,鉴此,完善实验报告和评审体制、提供实验细节原始记录、执行严格的论文评审标准就成为必需。在价值论上,学术不端行为(如P值篡改、择优选择、结果已知之后假设等)成为实验“可重复”困难的一个重要方面,必须杜绝。

**关键词** 生态学实验;可重复;本体论;认识论;方法论

对于生态学实验的“可重复”,国外学者作了区分,分别用“repeatability”、“reproducibility”和“replicability”表示。综合考察生态学相关文献对这3个单词的使用背景以及这3个单词自身的含义可发现,“repeatability”主要表示运用相同的实验仪器,根据相同的实验原理,依据相同的实验流程,进行相同的实验操作,以得到相同的实验现象乃至实验结果;“reproducibility”主要表示的是某一个实验结果的相同,而不必要求相同的实验原理、实验仪器、实验流程甚至实验现象;“replicability”主要表示的是基于随机的不同的实验单元所进行的多次实验的可重复。为了区别这3个单词,笔者分别称它

们为“可重现”、“可再现”和“可复现”。本文所用“可重复”如无特别说明,涵盖这3种。

“可重复原则”是科学实验的一个重要原则。一般来说,一项科学实验只有能够被自己或他人重复,该项实验的结果才可能被自己或他人接受并被认为是正确的,否则,该项实验就得不到科学共同体的承认。传统的科学实验通常是在实验室中进行的,实验者能够控制实验过程中所涉及的因素或必需的条件,如温度、光周期、培养基、样本的性别和年龄等,并且排除干扰,从而使得实验的可重复较强甚至很强。但是,生态学实验研究的是自然界中存在的生物与环境之间的关系,

收稿日期:2018-01-23;修回日期:2018-02-08

基金项目:国家社会科学基金重大项目(16ZDA112);教育部人文社会科学研究规划基金项目(13YJA720019)

作者简介:肖显静,教授,研究方向为生态学哲学及科学技术与环境论,电子信箱:xxjing@ucas.ac.cn

引用格式:肖显静.生态学实验“可重复”困难的原因及对策[J].科技导报,2018,36(6):8-16;doi:10.3981/j.issn.1000-7857.2018.06.001

不能像传统科学实验那样采取各种各样的手段和仪器设备去人工地建构对象和干涉对象,以获得对实验对象的各种各样的认识,而只能凭借自身独特的实验类型,在测量实验——“测量”自然、操纵实验——“处理”自然、宇宙实验——“模拟”自然、自然实验——“追随”自然的过程中,尽可能获得“自然状态下”生物与环境之间关系的认识<sup>[1]</sup>。也正因为这样,生态学实验更多地野外进行。由于野外各种环境因素变异性较大甚至非常大,难以较好地控制环境或实验条件<sup>[2]</sup>,再加上其他方面的原因,使得生态学实验的有效性、准确性和精确性受到影响,直接导致“重复”生态学实验变得艰难。生态学实验“可重复”较差甚至不能重复,成为笼罩在生态学研究头上的阴影,被生态学界内外所诟病。由此,需要高度关注生态学实验的“可重复”。

文献调研表明,国外对生态学实验“可重复”的研究有一些,但是比较零散,没有系统总结;国内尚未见这方面的研究。鉴此,有必要系统收集并分析国内外的相关文献,梳理出生态学实验“可重复”困难的原因,在此基础上,针对生态学实验研究现状,提出相应的改善对策。

## 1 生态学实验“可重复”困难的本体论考量

### 1.1 实验对象复杂性导致生态学实验“可重复”困难

#### 1.1.1 自然的时间空间变化性导致生态学实验“可重复”困难

生态系统中的生物和非生物,它们自身内在的关系、它们之间的相互作用和关系,以及它们与环境的关系,会随着时间和空间的变化而变化。这导致物种分布、种群分布以及群落结构和演替等出现较大的甚至很大的差异。对此,施尼策尔(Schnitzerand S A)等指出,这可能导致重现(repeatability)实验研究的结果与最初的研究结果不一样;即使我们能够复现(replicability)完全相同的野外实验,快速的环境变化可能导致生态的相互作用,它们的大小,甚至是方向,可能已经永久地改变了,这可能产生不同的结果<sup>[3]</sup>。

这与传统的实验室实验不同。传统的实验室实验虽然也存在自然的时间变异性——随机的或混乱的微小变化,并且可能会对不同时间所做的同一实验的结果造成较大影响,但是,在实验室条件下,这些变化可

以很好控制,并且使得它们很小,甚至可以忽略。在生态学实验中,相关因素随着时间和空间的变化而变化,这种变化可能一直很重要,也可能偶尔很重要;可能当前很重要,也可能将来很重要。这种时间和空间上大的自然变化性给生态学实验的重现或再现带来无法预料的困难,也使得生态学理论和实验不能得到重现(再现)检验<sup>[3]</sup>。即使在同一地点,在一个很短的时间和空间间隔内,要重现或再现一个生态学实验,也是比较困难的;在一个较长的时间间隔内,重现或再现一个实验就更加困难了,如持续的环境变化将使最近的研究与几十年前进行的对比研究复杂化。亚伦(Ellison A M)甚至认为,生态学现象是背景依赖的,而且背景会随着时间和空间的变化而变化,如此,精确地或量化地重现生态学野外实验,事实上是不可能的<sup>[4]</sup>。

#### 1.1.2 大尺度的限制导致生态学实验“可重复”困难

哈格罗夫(Hargrove W W)认为,在大尺度上进行操纵或设置复现(重现),并且维持长时间的操纵是很困难的。操纵往往会引起干扰,并通过其他系统组分传递意料不到的影响。设置了复现(重现)意味着该实验将是可被复现(重现)的,即某种程度上该实验是独立于时空的。但是,随着时空尺度的增大,设置复现(重现)的难度也增大了;除了费用等与人有关的原因,原来的奇异事件也会变得常见起来<sup>[5]</sup>。辛德勒(Schindler D W)认为,虽然“全生态系统”(whole ecosystem)实验得到的结果更真实,但要在大尺度“全生态系统”实验设置复现(重现)则几乎是不可能的。以湖泊为例,辛德勒发现虽然有46个湖泊可以用来做实验,但是仍然难以发现有近似复现(重现)的湖泊。造成此结果的原因除了昂贵的费用之外,还有湖泊之间的差异,包括动物区系、湖水更新时间(决定化学变化响应速率)以及化学物质的浓度等方面<sup>[6]</sup>。

即使不考虑上面这点,而假设可以进行“重现”实验,但是,一旦考虑到时间和空间的尺度较大,此类实验仍然不能顺利完成。施尼策尔(Schnitzerand S A)等指出:“重现生态学研究的另一个重要障碍是它们持续的大尺度的空间和时间。”<sup>[3]</sup>“在过去的40年中,一些最具影响力的生态学研究在空间和时间尺度上都非常大,重现许多这样的研究需要付出巨大的努力。对于某个人来说,甚至需要花费其整个职业生涯的几十年时间来重现几年前发表的有影响力的生态研究。”<sup>[3]</sup>

事实上,林肯(Likens G E)等关于生态系统的研究<sup>[7]</sup>,以及为了重现贯穿整个巴拿马热带森林,而在25年中的每个月收集木本植物物种物候学和生产率方面的数据,以用于判断是否复现(重现)<sup>[8]</sup>,就展现了这种困难。

而且,在大的空间和时间维度上进行的生态学野外实验按理说是最真实的,不过由于后勤和经济方面的限制,它们是最不可能被实验重现的。简单地说,谁能够大量地进行大规模的实验重现研究呢?谁会资助它们呢<sup>[9]</sup>?

## 1.2 因地制宜,提高生态学实验的“可重复”

### 1.2.1 使用易于处理的生物或系统来阐明过程

生态学野外实验的目的是研究生态现象并描述其过程,超越特定分类单元或系统的细节,得到一般过程的实验结果。鉴此,应该使用易于处理的“模型”生物或系统做实验。如对于“模型”生物,必须具有某些目标变量如密度、身体大小、繁殖力等,能够面对较大范围的自变量强度变化做出响应;必须在研究的空间和时间尺度上便于观察;必须以最不具侵入性的方式被操纵,因为操纵的侵入性越大,实验结果的人工性就越大,模型生物的有效性就越差。这样处理之后,生态学模型实验才更具“可重现性”。

### 1.2.2 选择同质性的或平衡的系统进行研究

对于大尺度的生态学实验,如区域生态学,为了提高它的“可重现”,往往会假定同质性和平衡。这些初始假定排除了区域生态学研究那些异质性的时空相互作用,从而呈现出一定规律性的、可重现的现象<sup>[9]</sup>。

### 1.2.3 进行微宇宙实验

由于微宇宙实验的尺度较小,通常是在瓶子之类的实验室器皿中进行,对环境条件可以进行良好的控制,而且可以对实验的处理设置复现,因此,它的可重现较高。阿博特(Abbott W)就认为,微宇宙实验的“可重现”已经成为证明该方法之于生态学研究正当有理的一个重要前提。阿博特发现,2篇意在重现微宇宙实验的文献声称其实验的优点是可重现和可控,但是结果发现这2篇论文的结果与其他论文的结果相冲突。在这种情况下,他建立了18个大圆玻璃瓶微宇宙进行实验,结果发现,在恰当的条件下,水生微宇宙显示了与统计试验中其他类型的复现单元类似的可复现。就统计学而言,这意味着可以建立成组的平行系统,在严格限定的实验中进行研究<sup>[10]</sup>。

## 2 生态学实验“可重复”困难的认识论考量

### 2.1 实验场所不能精确描述等造成生态学实验“可重复”困难

对于生态学实验,由于面对的对象常常是复杂的,由此导致对此认识的正确性受到影响,并进而影响到实验的“可重复”。如在生态学中,对于实验场所的精确描述将会促进生态学实验的“可重现”。然而,在野外,各种条件不断变化,难以得到很好的控制,进行一项可重现的实验变得更具挑战性。在尺度更大的一些地方,由于不可控变量持续增加,不确定的问题更为突出而且场所也越来越复杂<sup>[11]</sup>。这必然增加实验场所精确描述的困难,使得野外实验研究的可重现受到限制。

夏皮罗(Shapiro J T)和巴尔迪(Baldi A)从《Ecology》和《Oecologia》杂志中选取1年内发表的文献,计算了那些野外实验或者观察实验中与场所描述相关的文献数量,审察其中有关场所的描述,结果发现:在488篇文献中,27.3%的文献省略了地图和地理坐标,仅包含关于实验地区的模糊描述(例如仅提供县区或公园的名字);13.7%的文献包含地图;57.6%文献包含地理坐标。在包含空间坐标的那些文献的子集中,4.6%的文献仅仅精确到度,描述的地点可能包含了上千平方公里的地区;超过2/3的精确到了分,描述的地点覆盖了大约3.4 km<sup>2</sup>的地区;7.8%的文献精确到了秒,所涉及的地点缩小到几十平方米的地区<sup>[12]</sup>。

夏皮罗和巴尔迪将空间坐标与使用谷歌地图所获得的地点描述进行对比后发现,关于空间位置的错误是普遍存在的<sup>[12]</sup>。他们认为,发表的研究成果声称一组坐标落在某个区域范围之内(诸如郊外、岛屿、湖泊),而事实并非如此,表明该坐标应被视为“不准确”。而且,这种情况仅包含那些明显不准确的坐标。例如,那些声称是在岛上的坐标实际上是在临近的海洋中,或者是在其他与描述明显不同的郊外。如果坐标能够在所描述的地点范围内,比如在一个公园之内,就不会被归类于“不准确”,而不考虑这些坐标的精确度。他们发现,在这些发表的文献所提供的地理坐标中,接近16%的文献对于场所的描述都是不准确的<sup>[12]</sup>。

以上仅是从生态学实验场所精确描述的角度阐述实验可再现的困难。事实上,生态学期以来受到广泛的批判,有学者认为生态学中缺乏进步、生态学中没

有出现普遍性的理论、生态学的概念有缺陷、生态学家不能检验他们的理论等<sup>[13]</sup>。所有这些也可以造成生态学实验认识正确性的缺乏,从而导致其可重现或可再现的困难。

## 2.2 提高生态学实验认识的正确性,改善生态学实验的“可重复”

仍然以生态学实验的场所确定为例,夏尔罗和巴尔迪在指出“场所的确定和描述不准确从而导致生态学实验‘可重现’困难”之后,认为地理坐标可能不是描述研究场所的最好方法。他们建议使用标记语言文件(keyhole markup language, KML)(例如谷歌地图使用的)来储存信息和图片,这种文件可以被纳入大多数期刊的电子增刊和附录中<sup>[12]</sup>。

应该说,他们的这一建议还是可行和恰当的。地理坐标尽管可以对一个特定地点进行精确描述(例如一片林地或者一个小池塘),但是对于那些在更大区域范围进行的研究就不那么有用了(也就是说,坐标本身并不能对重现研究提供足够的信息)。不仅如此,在许多区域,谷歌地图拍摄的图片可以缩放得足够大以便使用者甚至能看清每一棵树,因而使用者也就能更好地了解地形,这也有助于研究者同时运用多重尺度法进行分析。

夏尔罗和巴尔迪进一步指出,可以通过以下方法尽可能精确地确定地点:通过GPS装置进行的野外测量,并且或者用谷歌地图或其他在线“电子探测器”获取图片;在地点坐标上提供经度和纬度<sup>[12]</sup>。

他们的上述观点是可行的。卫星图像和航空图片(诸如谷歌地图所拍摄的),能够显示坐标和地图所不能显示的地点的相关细节。与野外观测相联系,这些图片一是能够提供精确、详细的地理和植被覆盖度的相关信息,二是能够减少进行野外研究的费用,三是有助于长期的检测并允许研究者重现实验以说明景观尺度的变化。他们认为,精确的地点信息,特别是由标记语言文件所提供的信息,能够提高生态学研究的可重现和质量<sup>[12]</sup>。

除此之外,推进生态学认识的正确性、明确生态学概念的内涵和外延、增强生态学理论的普遍性、使生态学家能够更加充分地检验理论等,也是从认识论角度提高生态学实验“可重复”的有效路径。

## 3 生态学实验“可重复”困难的方法论考量

### 3.1 方法论细节缺乏导致生态学实验“可重复”困难

2001年,安德森(Anderson D R)等通过研究发现:“许多文献没有报告响应尺度或适当的误差线。”<sup>[14]</sup>2006年,菲德勒等的研究也支持了这一观点<sup>[15]</sup>。

2013年,兰德(Land M)等全局性地进行“从流经人工湿地的水流中去除总的氮和磷的定量研究”,以此揭示方法论细节上匮乏的状况。结论是:如果要进行有意义的综合,那么在121项研究中,就有67项研究在方法论细节和/或统计学设计上不充分<sup>[16]</sup>。2013年,海德威(Haddaway N R)等判决性评价现有的生态学研究证据,发现在已经发表的学术研究文献中,并没有提供足够的方法论细节以满足重现性实验研究。比如,一份关于废弃高海拔土地对环境和社会经济影响的系统性材料,在190项研究中有111项缺乏方法论细节的描述。更具体地说,38项研究没有报告调查时间,40项研究没有报告干预时间,28项研究没有描述复现的程度,105项研究没有描述实验发生的空间维度<sup>[17]</sup>。

2014年,海德威等对上述问题进一步展开研究,发表题为“低海拔的泥炭地进行有节制的开采:土地管理对温室气体和二氧化碳排放的影响的一个系统性评价”的论文,发现140项研究中有39项研究在方法论细节上较为匮乏。例如,没有说明管理活动的时间尺度及采样的具体时间或时间段,没有指出使用的复现次数,没有描述控制和恢复区域的相对位置<sup>[18]</sup>。2014年,瑟德斯特伦(Söderström B)等进行了“农田管理对土壤有机碳的影响”的系统评价研究,发现500项研究中有70项未能说明实验设计,如土地分块及随机配置等<sup>[19]</sup>。

2015年,海德威(Haddaway Neal R)和费尔赫芬(Verhoeven Jos T A)对农地管理对土壤有机碳的影响的系统研究发现,过去22年来,数据的遗失仍然是一个大问题,平均每年有13%的研究( $\pm 8.0$ [SD])没有报告样本参数。海德威和费尔赫芬通过批判性地评估大量重要证据后发现,方法论细节的欠缺已经成为生态学实验“可重现”的障碍<sup>[20]</sup>。

2017年,菲德勒等通过考察现有的研究以及他们自己的研究指出:“不幸的是,现有的证据表明,不完整的报告是非常普遍的。”<sup>[21]</sup>

上述方法论细节的缺乏,直接导致对此所进行的元分析以及“可重复”的困难。高森(Garssen A G)等就举例说明,缺乏信息的类似问题也同样发生在元分析中<sup>[22]</sup>。菲德勒等认为,对于补充材料或在线存储库中提供的数据,作者对它们的描述常常是错误的或不够充分,这导致再分析(re-analysis)的困难<sup>[21]</sup>。如吉尔伯特(Gilbert K J)等发现他们检查的60个分子生态学数据集中35%的情况都是这样<sup>[23]</sup>。可诺奇威(Koricheva J)等指出,不完整的报告对结果“可再现”造成了严重的障碍,限制了数据元分析的有效性,并可能使元分析和系统审查的过程弱化而产生偏差<sup>[24]</sup>。

### 3.2 完善实验报告,实现生态学实验的“可重复”

生态学实验方法论细节的缺乏,在不完整的实验报告中可以发现,典型地体现在方法的不透明上:没有或不能披露所有符合研究条件的样本量,以及确定这些样本量的方法;没有或不能描述用于选择研究主题和分配处理的理由及其方法;没有或不能展现数据丢失处理以及数据审查的方法;没有或不能给出与效应量或参量相对应的变量的充分细节;没有或不能明确所涉及的假设是先验的还是事后的等。鉴此,就要求生态学实验者在这些方面下大力气,以提供完整的生态学实验报告,提高生态学实验的“可重复”。

不过,考察现行的期刊指南、执行标准和激励措施,不足以完全且明确地要求实验者提供相关方法和分析报告。因此,西蒙斯(Simmons J P)等提出一个简单的解决方案,即要求所有作者对于方法部分,发表以下声明:“我们报告,在我们的研究中,我们是如何确定我们的样本量,是如何排除那些数据的(如果有的话),是如何进行那些操作的,又是如何使用那些测量的。”<sup>[25]</sup>菲德勒等也提出类似的对策——“报告登记”(“registered reports”)制度,具体而言就是期刊承诺论文的发表基于以下政策:论文的介绍、方法和计划的分析,以及对稿件的同行评审。在这一政策下,评审人员和编辑必须根据研究的基本原理(例如知道这个问题的答案有多重要)、方法(例如能否提供研究设计和分析回答问题的能力)做出决定,不能被结果是否重要所影响<sup>[21]</sup>。他们提到,在不同的学科,已经有30家期刊开始用某种形式实行“报告登记”制度。这为传统的同行评审过程提供了另一种选择<sup>[21]</sup>。

不仅如此,海德威等对作者、审稿人和期刊编辑提出了相应的要求。

对于生态学实验者或作者本人,基于他们在系统研究中的严格评价的经验,海德威等建议在实验和准实验研究的手稿中遵守以下最低要求:(1)实验设置:①户外研究:详细的研究地点(纬度和经度),有影响的气候条件;②实验室研究:受控条件(温度、光周期、有效的试剂);(2)研究日期和持续时间;(3)记录样本选择和治疗分配的选择过程(有目的的、随机的、限制的等);(4)真实的复现水平;(5)二次抽样的水平(量的重现实或者复现内采样(within-replicate sampling))(注:所有实验都要求实验处理有多个复现,目的在于定量非实验因素导致的变异。因此, $n$ 个实验因素,每个因素 $m$ 个水平的实验,就有 $n \times m$ 个处理,如果每个处理复现 $k$ 次,总共的实验小区就有在 $n \times m \times k$ 个。一般而言,在获取实验的目标参数时,要对所有小区进行采样,其中包括了每个实验处理的 $k$ 个复现采样,这些是真复现。考虑到小区比较大,或者小区内变异较大等因素,往往会在每个小区内采用多点取样,就是within-replicate sampling,通常的做法是把它们混合成一个样品来测定这个小区的参数。但是有些时候为了表征小区内的变异,可能对在一个小区内多个地点采集的样品不混合,而是分别测定目标参数,这样测得的值就是within-replicate sampling的值,这些不是真复现,而是复现内采样。within-replicate sampling可以翻译为“复现内采样”);(6)采样精度(复现内采样或复现);(7)研究空间尺度(复现的尺寸和研究区域的空间尺度);(8)研究设计(比如前-后(before-after)、控制-影响(control-impacts)、时间序列(time series)、前-后-控制-影响(before-after-control-impacts));(9)结果的测量方法和设备;(10)对所有的数据操作、建模或统计分析进行描述<sup>[20]</sup>。

对于审稿人或期刊编辑,海德威等认为,在主要研究缺少方法信息的地方存在着几个可能的解决方案:(1)联系相应的作者,要求获得信息;(2)检查同样的实验的相关底稿是否已发表,如发表,则就可以可靠地假定正在发表的论文的方法与此前的方法是相同的,从而提取此方法细节;(3)在元分析中进行敏感性分析(sensitivity analysis),以检验缺少重要的方法论信息对研究的影响(但也提供了足够的数量数据);(4)一旦发现,在专用数据库中发布丢失的信息[例如SRDR(<http://srdr.ahrq.gov>),或“已发布”的平台,比如PubPeer(<https://pubpeer.com>)],让未来的读者更容易找

到; (5)从长远来看,要促进提高报告标准:①改进当前期刊指南,提高生态学实验研究的报告标准<sup>[26]</sup>,建立对方法论细节的普遍授权;②加强高对评稿人员的指导,确保他们对方法论的重现性实验进行筛选;③提高人们对重现性实验的重要性的认识,要特别重视二次综合(secondary synthesis)以及它们对实验合法性的帮助和影响。选项(1)、(2)、(3)是广泛适用的,只需最小的付出就可以实现。然而,考虑到机构之间的研究人员的流动,通讯地址的变化,预计3年至5年之后原先的电子邮件的回复率可能会特别低。选项(4)和(5)可能需要相当大的努力,并且需要科学界的集体努力<sup>[20]</sup>。

## 4 生态学实验“可重复”困难的价值论考量

### 4.1 “学术不端”导致生态学实验“可重复”困难

一项重要的研究发现,科学实验缺乏可重现的一个可能原因是数据造假和其他的不道德行为,如选择性地删除一些数据点等。尽管我们主观上期望这种行为在生态学中并不常见,但是,在其他科学领域大量的论文因此被驳回使我们对上述主观期望存疑。事实上,发表高影响力的论文,既是成功竞聘工作的必要条件,也是支撑生态学研究的需要,在此压力下,仍然相信生态学中没有多少数据造假等类似行为,未免显得天真<sup>[27]</sup>。

这就是说,“不发表,就灭亡”(publish or perish)的文化,是造成生态学实验“不端行为”的一个重要原因。只有那些发表出来的成果,才会被当作成果,也才会被人们知晓,并进而成为科研人员晋升、提职等的根

据。这样一来,科研人员就要千方百计地、更多地发表论文。对于那些失败了或者在统计学上呈现非显著性的生态学实验成果,按照实验研究成果发表的规则,就只能报废或放在文件抽屉里尘封起来,成为废物。在这种情况下,某些科研人员出于对“不发表,就灭亡”文化的尊崇和恐惧,通过所谓的“研究人员的自由度”(“researcher degrees of freedom”),对那些不符合发表要求的、不合格的研究论文,人为地使之“成功”或者“恢复”统计的显著性,使之“复活”,以最终达到发表的目的<sup>[28]</sup>。

这就是众所周知的“成问题的研究实践”(questionable research practices,简称 QRPs)。菲德勒等关注零假设统计检验(null hypothesis statistical testing, NHST)相关的“可再现”议题,通过对早期杂志《Ecology》、《Journal of Ecology》、《Biological Conservation》和《Conservation Biology》等有关“零假设的显著性检验”(null hypothesis significance testing,简称 NHST)的调查报告的重新检查,发现 QRPs 出现的情况并没有减弱的迹象:在2005年,84%的文献(200篇文章中的167篇)报道了 $P$ 值;在2010年,对应的数字为90%(170篇文章中的153篇)报道了 $P$ 值<sup>[15]</sup>。

菲德勒等对此展开进一步的系统研究,认为其形式主要包括 $P$ 值篡改、择优选择、结果已知之后假设(hypothesizing after the results are known, HARKing)等(表1)<sup>[21,29]</sup>。

不仅如此,菲德勒等提出,一些 QRPs,只存在于概率统计显著性实验中,但是在其他范式中,即在统计学显著性检验之外,这样的对应物也存在<sup>[21]</sup>。

表1 成问题的研究实践,在文献中夸大假阳性率,导致更少的可再现研究

Table 1 Problematic research practice, exaggerated false positive rate in the literature, leading to less reproducible research

$P$ 值篡改	在决定是否收集更多数据之前,检查统计结果的统计显著性 提前停止数据收集,因为结果已经达到统计上的显著性水平 只在统计意义的影响预兆之后而不是在报告数据排除的影响之后,决定是否排除数据点(例如,离群值) 舍入 $P$ 值以达到统计学显著性阈值(例如,将 $P < 0.053$ 作为 $P < 0.05$ )
择优选择	未能报告没有达到统计显著性或其他阈值的依赖变量或响应变量或关系 未能报告没有达到统计显著性或其他阈值的条件或处理办法
结果已知之后假设	呈现出一个事后调查结果,好像它一直是假设的

### 4.2 遵守学术道德,遏制“成问题的研究实践”

影响生态学实验“可重现”的一个重要原因是“不发表即死亡”(Publish or Perish)的文化,这导致“成问

题的研究实践”(QRPs)在生态学实验中有所抬头,从而直接影响到生态学实验的“可重现”。鉴此,应该对生态学研究进行诚实性调查,确实弄清此项文化在他

们的QRPs中的影响程度,即是否确实是“不发表即死亡”的盛行导致QRPs的程度加剧,从而导致生态学实验“可再现性”的降低,这是其一。

其二,菲德勒等提出,可以建立预先注册数据库(preregistration databases),作为注册报告的先驱或替代。这一数据库也是资料库,其中,研究人员在数据收集与处理之前,公开对研究的问题、假设或预期、方法和计划好的分析作出承诺(<https://cos.io/prereg>)。预先注册的条例可以广泛应用于各种各样的研究,而不仅仅是那些关联假设检验的研究。另外,作为一种及时遏制QRPs的策略,该条例已经在其他学科中被大力提倡<sup>[21]</sup>。

其三,改变“不发表即死亡”的文化,不以文章数量来评价,而以文章质量以及真实性来评价。施尼策尔等就说,尽管伪造数据仍然很难被发现,但是增加出版生态学数据作为论文的补充,有助于减少欺诈。而且从根本上讲,任何学术领域的成功都取决于其公民的诚实和道德,我们希望这在生态学中能保持下去<sup>[3]</sup>。

## 5 结论与讨论

### 5.1 结论

1) 生态学实验对象自身的时间、空间的变异性以及大尺度的限制,是其“可重复”困难的根本原因。要克服此类困难,一是使用易于处理的生物或系统来阐明过程,二是选择那些同质性的或平衡的系统进行研究,三是进行微宇宙实验。

2) 由于人类认识能力的阶段性和有限性,也由于生态学实验认识对象的复杂性,导致生态学实验认识往往具有不确定性,缺乏准确性和精确性,由此直接影响到实验的“可重复”。从目前看,不能精确描述实验场所,是这方面的一个表现,改善之路是运用最新的定位系统。

3) 完整的生态学实验报告对于实验的“可重复”至关重要。从目前看,方法论细节缺乏在生态学实验中普遍存在。鉴此,一是要求实验者提交完整的实验报告以及方法论细节;二是建立新的明确生态学实验“可重复”的论文发表的相关制度,并且促使实验者、评审者和编辑按照这些规定执行;三是对生态学研究进行培训和指导,使他们意识到方法论细节对于生态学实验“可重复”的重要性。

4) 出于论文发表的压力或者其他原因,在生态学

实验中也存在“学术不端问题”,典型地表现在“P值篡改”、“择优选择”以及“结果已知之后假设”方面。为此,对于科学共同体,需要改变“不发表,就死亡”的文化,实施“道德承诺”和相应的“道德失范处罚”制度。

### 5.2 讨论

1) 对于生态学实验的“可重复”应该客观地、理性地对待。对于某些生态学实验,由于其对象及其过程的复杂性,导致其本身不具有“可再现”甚至“可重现”,此时对此进行“可重复”实验就是不合理的了。而且,对于某些规模巨大且异常复杂的生态学实验,即使“可重复”现实可行,由于所需人力、物力、财力巨大,是否有必要对此进行“可重复”实验,也需要考证。

2) 对于那些“不可重复”或者没有必要“重复”的实验,不能放任自流,而应该对此进行“元分析”(meta-analysis)或元研究(meta-research),即对该项实验研究本身进行研究,以明了其是否真的具有“不可重复”或没有必要“重复”,“不可重复”或者没有必要“重复”的程度和性质如何,“不可重复”或者没有必要“重复”的原因以及合理性怎样,应该如何保证“不可重复”或者没有必要“重复”的生态学实验的正确性等。考虑到这点,菲德勒等指出,“元分析”特别适用于这样一些领域——直接的或彻底的复现(重现或再现、复现)并不可行的领域,以及那些甚至并不依赖实验数据的领域<sup>[21]</sup>。

3) 对于那些现实进行的“可重复”实验后所表明的“可重复”、“不可重复”或者“低重复”生态学实验,也要进行“元分析”或“元研究”。因为,对于这样一些实验,它们的“可重复”可能是以对实验对象的复杂性过分简化为代价的,它们的“不可重复”或者“低重复”可能是由认识的不正确或不精确造成,或者是由相应的方法论细节缺乏造成,或者甚至就是由某些学术不端造成。对于一个生态学实验,现实中所进行的“可重复”验证或者“不可重复”展现并不能充分说明这样的实验的合理性,对于生态学实验,“元分析”或者“元研究”必不可少,由此一定程度上可以明了一个生态学实验是否真的具有“可重复”,以及这样的“可重复”是否合理。

4) 对于本体论、认识论、方法论意义上的生态学实验“可重复”困难,应该采取各种措施加以改善,不过,这样的改善不应以损害或者严重损害生态学实验的实在性或者逼真性为代价。一个为了“重复”而“重复”的生态学实验不是我们需要的,我们需要的应该是在“可重复”与“实在性”或者“逼真性”间达到权衡的生态学

实验。关于这点,当另文阐述。不仅如此,本文所论及的认识论、方法论意义上的生态学实验“可重复”是就“重现”的意义而言的,事实上,“再现”一个实验,也就是“重复”一个实验结果,一定意义上甚至在很大程度上可以脱离原先实验细节乃至是否正确的束缚。

5) 对于本文生态学实验“可重复”困难原因分析及其改善措施的提出,是必要的,但不是充分的,应该还有造成“可重复”困难的其他方面的原因以及相应的对策,如贝克(Baker M)通过对1576名科研人员的问卷调查发现,导致实验无法重复的原因有多种,即选择性报道、论文发表的压力、统计效力低、在原实验室就不充分、监管以及指导不足、缺少实验技术和参数、实验设计太差、无法获得原始数据、作弊造假、同行评议不够<sup>[30]</sup>。就此而言,本文的工作还有待于扩展和深化。而且,本文没有对国内生态学实验“可重复”状况进行调查,也没有对造成这种状况的原因等作进一步研究,这些应该是生态学研究重要且迫切需要研究的内容。

### 参考文献(References)

- [1] 肖显静. 生态学实验的“自然性”特征分析[J]. 自然辩证法通讯, 2018, 41(3): 1-9.  
Xiao Xianjing. The characteristics of "naturalness" in ecological experiments[J]. Journal of Dialectics of Nature, 2018, 41(3): 1-9.
- [2] Polis G A, Wise D H, Hurd S D, et al. The interplay between natural history and field experimentation[C]//Reseratis W J, Bernardo J. Experimental Ecology: Issues and Perspectives. Oxford:Oxford University Press, 1998: 257-258.
- [3] Schnitzerand S A, Carson W P. Would ecology fail the repeatability test[J]. BioScience, 2016, 66(2): 98-99.
- [4] Ellison A M. Repeatability and transparency in ecological research[J]. Ecology, 2010, 91(9): 2536-2539.
- [5] Hargrove W W, Pickering J. Pseudoreplication:a sine qua non for regional ecology[J]. Landscape Ecology, 1992, 6(4): 251-258.
- [6] Schindler D W. Replication versus realism:the need for ecosystem-scale experiments[J]. Ecosystems, 1998(1): 323-334.
- [7] Likens G E, Bormann F H, Pierce R S, et al. Biogeochemistry of a forested ecosystem[M]. New York: Springer-Verlag, 1977.
- [8] Wright S J, Calderon O, Hernandez A, et al. Are lianas increasing in importancein tropical forests: A 17-year record from Panama[J]. Ecology, 2004, 85(2): 484-489.
- [9] Risser P G, Karr J R, Forman R T T. Landscape ecology: Directions and approaches[M]. Champaign: Illinois Natural History Survey Survey Special Publication No.2, 1984.
- [10] Abbott W. Microcosm studies on estuarine waters I: The replicability of microcosms[J]. Water Pollution Control Federation, 1966, 38(2): 258-270.
- [11] Kraufvelin P. Baltic hard bottommesocosms unplugged: Replicability, repeatability and ecological realismexamined by non-parametric multi-variate techniques[J]. Journal of Experimental Marine Biology & Ecology, 1999, 240(2): 229-258.
- [12] Shapiro J T, Báldi A. Lost locations and the(ir)repeatability of ecological studies[J]. Frontiers in Ecology and the Environment, 2012, 10(5): 235-236.
- [13] 福特. 生态学研究的科学方法[M]. 肖显静, 林祥磊, 译. 北京: 中国环境科学出版社, 2012: 438-444.  
Ford. Scientific method for ecological research[M]. Xiao Xianjing, Lin Xianglei, trans. Beijing: China Environmental Science Press, 2012: 438-444.
- [14] Anderson D R, Link W A, Johnson D H, et al. Suggestions for presenting the results of data analysis[J]. Journal of Wildlife Management, 2001, 65: 373-378.
- [15] Fidler F, Burgman M, Cumming G, et al. Impact of criticism of null-hypothesis significance testing on statisticalreporting practices in conservation biology[J]. Conservation Biology, 2006, 20: 1539-1544.
- [16] Land M, Graneli W, Grimvall A, et al. How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal: A systematic review protocol[J]. Environmental Evidence, 2013, 2(1): 16.
- [17] Haddaway N R, Styles D, Pullin A S. Evidence on the environmental impacts of farm land abandonment in high altitude/mountain regions: A systematic map[J]. Environmental Evidence, 2013, 2: 18.
- [18] Haddaway N R, Burden A, Evans C D, et al. Evaluating effects of land management on greenhouse gas fluxes and carbon balances in boreo-temperate lowland peatland systems[J]. Environmental Evidence, 2014, 3: 5.
- [19] Söderström B, Hedlund K, Jackson L E, et al. What are the effects of agricultural management on soil organic carbon (SOC) stocks[J]. Environmental Evidence, 2014, 3: 2.
- [20] Haddaway N R, Verhoeven Jos T A. Poor methodological detail precludes experimental repeatability an hampers synthesis in ecology[J]. Ecology and Evolution, 2015, 5(19): 4451-4454.
- [21] Fidler F, Chee Y E, Wintle B C, et al. Metaresearch for evaluating reproducibility in ecology and evolution[J]. BioScience, 2017, 67(3): 282-289.
- [22] Garssen A G, Verhoeven J T A, Soons M B. Effects of climate-induced increases in summer drought on riparian plant species:a meta-analysis[J]. Freshwater Biology, 2014, 59(5): 1052-1063.

- [23] Gilbert K J, Andrew R L, Dan G B, et al. Recommendations for utilizing and reporting population genetic analyses: The reproducibility of genetic clustering using the program structure [J]. *Molecular Ecology*, 2012, 21(20): 4925–4930.
- [24] Koricheva J, Gurevitch J, Mengersen K. *Handbook of meta-analysis in ecology and evolution*[M]. Princeton: Princeton University Press, 2013.
- [25] Simmons J P, Nelson L D, Simonsohn U. A 21 word solution [EB/OL]. (2016–11–11). <http://dx.doi.org/10.2139/ssrn.2160588>.
- [26] Hillebrand H, Gurevitch J. Reporting standards in experimental studies[J]. *Ecology Letters*, 2013, 16(12): 1419–1420.
- [27] Schnitzer S A, Carson W P. Would ecology fail the repeatability test[J]. *BioScience*, 2016, 66(2): 99.
- [28] Simmons J P, Nelson L D, Simonsohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant[J]. *Psychological Science*, 2011, 22: 1359–1366.
- [29] John L K, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling[J]. *Psychological Science*, 2012, 23: 524–532.
- [30] Baker M. Is there a reproducibility crisis[J]. *Nature*, 2016, 533(7604): 452–454.

## The "repeatability" difficulty in ecological experiment

XIAO Xianjing

School of Public Administration, South China Normal University, Guangzhou 510006, China

**Abstract** The “repeatability” difficulty in the ecological experiment is an important issue. From the point of view of the ontology, mainly due to the natural variability, and partly due to the large scale, it is easy to use an organism system to illustrate the process, or to choose the homogeneity and balancing system to simulate the natural micro universe experiments. However, all these simplifications of the natural selection, or simplified simulations of the nature, although valid for a moment, are not always feasible. From the point of view of the epistemology, the ecology of the complexity of the subjects, the organic wholeness and historic nature make the validity of the understanding restricted, which directly affects the experimental "repeatability". Therefore, it is necessary to accurately determine the experiment site, and to clearly define the relevant concepts, to achieve the correctness(reality) of the ecology experiment and create a "repeatable" win-win situation. Methodologically, the incomplete experimental reports and the lack of relevant method-specific details are important causes of the "repeatability" difficulty of ecological experiments, thus it is necessary to improve the experimental report and the review system, to provide the original records of experimental details, to apply strict evaluation criteria. From the point of view of the axiology, academic misconducts, such as the *P*-hacking, the cherry-picking, and the hypothesizing after results(HARKing), etc., become important aspects of the experiment's "repeatability" difficulty and must be eradicated completely.

**Keywords** ecology experiment; repeatability; ontology; epistemology; methodology ●



(责任编辑 陈广仁)