

数据哲学构建的初步探析^{*}

刘 红 胡新和 (中国科学院大学人文学院 北京 100049)

[中图分类号]N031 [文献标识码]A [文章编号]1002-8562(2012)12-0082-07

通过对大规模数据的研究来探索自然规律,已经成为当代科学研究的特征之一,这不仅是对宇观世界、微观世界和复杂世界研究手段的客观要求,也得益于科学仪器包括计算机的高度发达和广泛使用。现代社会已经进入信息化和知识化时代,其最大的标志是数字化,2011 年全球数字化信息总量已超过 1.8ZB,5 年内增长了 9 倍。IDC 预测,未来 10 年总量将增长 75 倍。^[1]同时大规模数据的产生也催生了以大数据为研究对象,通过数据挖掘寻求科学规律的新的学科领域,这些学科领域对当代科学及社会活动都产生着越来越大的影响。“科学就是数据,数据就是科学”,“数据推动着科学的发展”,整个社会的经济、政治、环境和健康事业的发展,也需要数据的支持。^[2]数据(信息)与能源、材料一起被认为是三大战略资源。2012 年,美国奥巴马政府投资 2 亿美元专门从事大数据研究及发展计划。^[3]

数据为什么会有如此大的威力?它的本质是什么?有哪些特征?这些数据的哲学问题已经引起学者们的关注,近年来,有关数据研究的文献逐渐增多。尤其是在《综合》(Synthese)、《科学工程伦理》(Science Engineering Ethics)、《科学》(Sciences)等著名国际期刊上相继开设了专刊、专栏,论述数据所涉及的哲学、管理、共享等问题。本文将对数据的演变及其对科学研究范式的影响、学者们对其已有研究的状况进行梳理,并提出构建数据哲学的初步设想,尝试对数据哲学的定义、研究内容和方法进行探讨。

一 数据的内涵及其演变

数据来源于数、量等基本概念。数的概念源于远古时期的数觉,是人类对周围事物的多少、大小的认识。经研究发现,数觉并非人类所独有,数觉是人类祖先对世界的朦胧认识。数源于希腊语“Ἀριθμός”,古希腊时期的毕达格拉斯学派对数的神秘性进行了理性探索,推动了数的进一步发展。在 13 世纪到 14 世纪,在欧洲产生了数(number)、数字(amount)、数值(quantity)、量(amount)等术语的英语表达并得到广泛使用。此时,由于丈量土地、贸易以及技术发展的需要,数或量的概念中已经包含测量、观察等人类主动参与的、有意识地区分和鉴别事物的含义。

与数、量相比,“数据(data)”一词的内涵发生了根本性的变化,它不仅代表事物的多少、大小,用于鉴别事物,而且已经成为人类生产和生活的基础、依据,尤其是对近代科学的研究范式的形成产生了巨大影响。17 世纪 40 年代,产生了数据的英语表达“data”,其词源为拉丁语“dare”的过去分词“datum”,“data”是“datum”的复数形式。在拉丁语中,“dare”是一个非常重要的词源,表示与人无关的增加、添加,具有寄予、给予、赐给(to give)、“给出物”(something given)的意思。“data”对应的希腊语“δεδωμένο”也是给予(give)的意思。^[4]到了现代,在讨论几何学、数学、工程学和其他

* 本文系作者主持的中国科学院规划与决策科技支持系统建设“科技与经济社会互动系统——科研道德失范的研究”(GH11048)的阶段性成果之一。

学科问题时,术语“给予”(given)与“数据”常常可交换使用。总之,数据被赋予了具有相对客观性的“给予”和“依据”的内涵。

到目前为止,对数据一词尚无统一的定义,普遍观点认为数据用于表征事实、数字 (figure) 和观点。《现代汉语词典》将“数据”解释为进行各种统计、计算、科学研究或技术设计所依据的数值。^[5]在维基百科中^[6],数据被定义为关于事件的一组离散且客观的事实描述,是一个变量或一组变量的质和量的属性,是以图形、声音、文字、数、字符和符号等为基础的、典型的测量结果或者是一系列变量的观察结果。数据也被称为资料,是指描述事物的符号记录,可定义有意义的实体,涉及到事物的存在形式;在知识层次理论中,数据被认为处于最底层或贯穿于各层次之中,从它可抽象出信息,并派生出知识,最后形成人类的智慧,因此,数据被认为是构成信息和知识的原始材料。

科学技术的发展更是依赖于数据,数据已经成为观察现象的主要表征方式,成为各种统计、计算、科学研究、工程实施、技术设计、政策制定,囊括社会、经济、军事、教育、科研等活动的基础和依据。数据是电子计算机工作的基本“原料”。1946年第一次将“data”表示为“传输和存储计算机信息”。计算机术语“数据处理”(data processing)一词从1954年开始使用。^[7]

20世纪后期,由于计算机的使用,获得的数据在数量、类型、速度等方面均达到一个前所未有的新阶段,数据的表现形式也从传统意义上的数值扩展到了包含了文字、图、表等大量可以通过计算机转化为数字化形式的信息资源。近年来,在强大的技术支持下,海量数据的获取使人们意识到数据对科学研究内容和方法将产生巨大的影响。科学家们惊叹今天的数字时代已经进入到“数据丰富,而理论(或信息)匮乏”的阶段。人们迫切要去解读数据内部蕴含的规律。

二 数据与科学研究范式

从近代科学诞生至今,数据在科学研究范式中的地位和作用发生了根本性的转变。

1. 数据是近代科学研究范式的基础

与早期的哲学思辨相比,近代科学研究范式为数据提供了一席之地。在近代科学的产生和发展中,数据被认为是研究对象内在变化规律的外在表征,成为建立科学理论的基础;通过观察和记录从实验中获取的数据,应用逻辑推理形成科学理论,再经由重复的实验数据来验证和证伪科学理论的正确性,这种自然哲学的认识论基础和方法论路径在近代科学形成和发展中占据了主要地位。

各学科领域,如17世纪发展的物理学、18世纪真正意义上的化学革命、19世纪的生命科学的建立等,都经历了由定性的描述研究方法向以定量研究为基础的数理研究方法的转变,各学科的研究纲领在数理研究方法统领之下逐渐建立和完善起来。实验科学和理论科学成为近代科学的两大组成要素,数值计算是在理论科学和实验科学中普遍存在的科学研究方法。数据贯穿于实验科学和理论科学之中,在实验设计、结果分析、理论建立、理论判据等方面占据了不可替代的地位,成为近代科学研究范式的基础。

2. 数值计算成为现代科学研究范式的组成部分

在数值计算成为一个学科分支之前,计算仅使用人力或简单的计算工具,其功能和效率极其有限。20世纪中叶,引领第三次技术革命的电子计算机技术导致了数据处理方式的巨大变革,电子计算机不仅加快了运算速度,而且极大地增加了数据的处理量,还能通过数值模拟取代人的部分智能活动。

20世纪80年代,在物理学领域逐步形成了一门新兴的边缘学科——计算物理学,它是以电子计算机为工具、采用数学方法解决物理问题的应用科学,是物理、数学和计算机三者相结合的产物。以理论物理提供的理论原理和数学方法进行数学建模或以实验物理提供的大量实验数据为基础,通过计算机的模拟,解决物理学、力学、天文学和工程中复杂的多因素相互作用的问题,模拟结果由理论物理、实验进行检验和分析。计算物理学在20世纪80年代还只被作为沟通理论物理学与实验物理学之间的桥梁,

最近几年,随着计算机技术的飞速发展和计算方法的不断完善,计算物理学成为一门独立学科,与理论物理学和实验物理学一起成为现代物理学的三大支柱。

与计算物理学相似,利用先进仪器和计算机技术收集、处理数据,运用有关学科知识和数学知识进行数学建模,并有效使用数字计算机求解相关问题近似解的方法与过程,已经逐渐被应用到其他传统学科之中,形成如计算化学、计算力学、计算经济学等新型学科。以数据为基础的计算科学成为一门独立的学科,与实验科学、理论科学一并构成新的科学研究范式。

3. 数据共享

随着数据量的迅猛增加,数据的使用问题更为突出,对已有数据的共享成为人们关注的话题。高尔顿(Francis Galton)是生物统计学的开创人,1901年出版的《生物统计学》介绍了统计学在处理原始数据中的方法和意义,并强调了原始数据共享的理念,成为首位提出原始数据共享的科学家。^[8]在社会科学领域,从20世纪60年代早期起,许多专门讨论通过开发获取基本的社会科学数据的价值和好处的书籍和文章已经出现,他们探索提供这种路径的方法。许多文献强调由于用于社会科学研究的技术发生变化,从而对于社会科学数据的获取有直接的影响。^[9]1966年,国际科技数据委员会(CODATA)成立,旨在促进全球科技数据的共享。1985年出版的《数据共享》^[10]一书,对数据共享的主体、作用、方式进行了全面的分析和阐述。

在如何使数据最大化共享的同时,一些数据也需要保密,尤其涉及个人隐私时。科研机构、政府组织对科研人员、医生、护士应有有关数据的培训。教育的内容应该涉及数据的所有领域。科学家需要与不同领域的科学家交流,需要具备数据处理的相关知识,但是,目前对科学家如何进行科学数据教育还不清楚。^[11]乔思(Margi Joshi)和克拉格(Sharon S. Krag)^[12]提出在数据获取过程中,尤其是面向人体实验和动物实验,恰当的认证机构的认同是非常必要的,数据的恰当处理和保存是研究人员应予以考虑的问题。

4. 数据科学的诞生

20世纪末,在生物学领域开始的基因组测序技术飞速发展,积累了大量的生物学数据,如何理解这些数据,成为一种新的挑战。相同的数据问题也蔓延到各个学科领域,大到宇观的天文学研究,小到微观的基本粒子研究;以及复杂系统的研究,如气象学研究、社会学研究。中国学者朱扬勇在其《数据科学》^[12]中,将自然界的概念进行了扩展,提出数据自然界的概念。他认为人类在认识由宇宙和生命组成的真实自然界(real nature)的过程中,产生的成果存储在计算机系统中,在不知不觉中创造了一个由计算机中的数据构成的数据自然界(data nature),数据自然界中的数据以自然方式增长而不为人类所控制,数据自然界具有未知性、多样性和复杂性的特点。以大量观察和观测数据为研究对象,通过挖掘、提取等手段,寻求研究对象的内在规律的学科——数据科学应运而生。这一新学科是否叫数据科学目前还有争论,有的学者如李国杰院士认为应该为数据工程、数据技术,但无论如何,以数据为对象的相关研究已经得到学术界的关注。2001年,以CODATA创建的学术期刊《数据科学杂志》(CODATA Data Science Journal)为标志,科学技术数据科学作为一门独立的学科正式诞生。

科学技术数据科学的产生再次推动了科学的研究范式的变革。以电子计算机为技术手段,通过对海量数据的挖掘获得有用的知识,与传统科学的发展并驾齐驱,在不同学科领域催生一些新的研究方向,如地理信息科学、生物信息科学等如雨后春笋般出现。

三 数据的哲学认识现状

1. 数据的科学哲学研究

(1) 数据与现象

科学哲学中,科学实践由理论和观察构成的二元论是人们普遍接受的观点。1988年波根(J. Bogen)和伍德沃德(J. Woodward)^[13]将现象加入科学实践的二元论之中,构成理论-观察-现象的三元论,并首

次对数据和现象做了区别,其目的是为了帮助解释科学实践,特别是现象与理论之间的关系。他们的主要观点表现在如下几个方面:(1)在证明现象的存在时,数据起到重要作用,在大多数情况下,数据可以直接观察到,数据是现象的证据。(2)提出了数据现象推论,即只有通过数据作为证据才能显示现象的存在。(3)认为数据和现象的区别和联系在于,一方面,数据的典型特征是不能被理论预测和系统解释的;相对应的,良好的科学理论可以对有关现象进行预测和解释。现象是利用数据而发现的,但在大多数情况下,现象不能用有意义的可观察量术语进行描述;另一方面,数据是通过仪器记录的观察和测量的结果,是受人类的感官系统影响,获得的数据会因为外界干扰因素、统计推断、数据压缩等产生误差,但是如果能消除这些影响因素,就能产生可靠的数据。最后,他们得出的结论是,在不同的语境和测量观察方法下,数据是不同的,它具有独特性;而现象是自然固有的,其特性是稳定不变的,但是它是不易被观察到的,只有通过数据作为证据才能显示现象的存在。

波根和伍德沃德的数据与现象的概念自1988年提出以来,受到学术界关注,有支持也有反对。针对波根和伍德沃德认为现象和数据的区别既是有效的又是令人信服的观点,麦卡利斯特(McAllister)^[14]首先提出质疑,他认为他们没有给出数据与现象区别的恰当路径;同时,按照正确的定义,相对于世界特征而言,现象自身就是调查者,并提出现象的理论渗透观点。格利莫(Glymour)^[15]则批评波根和伍德沃德以及麦卡利斯特过于强调区分数据与现象的必要性,他认为通过经验推演的理论不存在像麦卡利斯特强调的理论渗透。

2008年9月,在海德堡举办了有关“数据·现象·理论:什么是有利于科学现象的概念?”国际研讨会,2011年,《综合》(Synthese)杂志发表了这次会议收集的论文专刊。波根和伍德沃德对数据与现象区别的概念成为中心话题,学者们从现象学、认识论和方法的关系^[16];数据的可靠性^[17];现象的稳定性问题^[18];经验等效性和数据等效性^[19];参数模型^[20]等方面进行了研究,其中拯救现象^[21]是波根对他们先前观点进行的补充说明和修正。

学者马查默(P. Machamer)^[22]认为波根和伍德沃德假设现象是由实验获得的数据构成。罗维(B. Lowe)和莫勒(T. Muller)^[23]认为波根和伍德沃德的数据/现象讨论,仅限于科学哲学的内部,他们力图将论证数据/现象区别更广泛地应用于哲学的模式处理之中。弗兰肯鲍格(B. Falkenburg)^[24]认为按照对科学实在论的不同视角,对于物理现象将有不同的认识。

除专辑之外,也有学者^[25]已经以波根和伍德沃德的观点为出发点,讨论了科学的自主模式和科学实在论。除上述关于数据与现象的关系研究之外,通过经验数据获得理论的科学实证论、精确性和决定性问题^[26]也是哲学家们关注的话题。

(2) 数据与理论

逻辑实证主义和证伪主义均将理论和数据之间的关系置于科学的核心。这是加拿大哲学家西斯蒙多^[27]对科学图景和技术图景的解读:

在解释科学活动的条理性问题时,逻辑实证主义认为通过将单个数据点转化为普通陈述,科学理论便得到发展。建立科学理论是一个推动过程。他们力图发展一种科学逻辑,以便使从个别事实到普遍主张的归纳过程更加牢靠。比如对数据进行无歧视的使用。

数据无法直接体现抽象的理论,因而理论和观察没有系统的联系。我们相信一个科学理论的理由,是基于此理论与数据之间的稳固联系。证伪主义具有怀疑精神,认为最好把进步看成是理论的不断精致化和扩展,以涵盖越来越多的数据。

逻辑实证主义和证伪主义都认为,科学之所以成其为科学,其特征在于理论和数据之间具有形式关系,不管是基于经验数据对理论大厦的理论建构,还是基于经验数据对理论的理性拒绝。实在论认为,大多数科学理论近似为真,一方面源于不断提高精确度的更好的数据,另一方面,无论是建构理论还是

证伪理论,都是为了接受更接近真理的理论。功能主义认为,规范是支配科学行为和态度的规则。理论由经验证据不充分决定时,简单性和精致性标准成为规范。

2. 数据的伦理学研究

数据的伪造和篡改等科学不端行为直接威胁着科研活动的正常运行,产生不端行为的根源和能否规避不端行为,成为目前学术界广泛关注的问题。

学者们在研究影响获取数据的外部条件及数据的性质^[28]的同时,对由于高度发展的各种技术对数据的处理和保存有着极其深远的影响,并可能带来的一些伦理上的问题给予了高度关注,克罗梅(Douglas Cromeey)和沃尔默(Benos and Vollmer)^[29]认为新的数据图像技术已经使数据处理变得更加灵活,这种灵活性对数据管理提出了新的伦理挑战。应该在可接受的行为和不可接受的行为之间有一个可以划分的界限。他认为大量地使用图像处理虽然不是有意欺骗,但是可能导致实际获得数据的误解。

3. 数据的价值

数据是科学仪器和科学家认识活动的产物,具有主观性特征。数据的准确性和精确性,是科学研究的基础,也是其价值所在。数据价值由获取数据的技术价值、应用价值和共享价值等方面构成。

数据的价值最突出的表现形式是其生命周期。数据的生命周期不仅仅与数据自身的使用有关,而且与科学家的认识、数据的管理等外在因素关系密切。在新技术不断涌现的时候,如何认识数据的生命周期的问题显得尤为突出。德国的JADE实验数据的再开发利用,^[30]使科学家们意识到在实验装置更新换代时,原有旧装置获取数据的价值并非相应消失。与科学范式的不可通约性形成鲜明对照,数据的可通约性是其价值的另一个表现形式,即数据的共享,将天文学与癌症研究方法相联系的桥梁正是数据。^[31]提高数据的应用价值是科学界不断追求的目标。气象科学界正在寻求一种开放的、友好的数据访问范式,以应对数据的复杂性与数据数量迅猛增加的局面,确保社会可以降低气候变化引发的脆弱性。^[32]为了确保数据价值的充分利用,对数据实施保护相当重要,数据的保护包括三个核心要素:保密性、完整性、有效性。^[33]

总之,从目前的研究文献可以看出,对数据的哲学研究尚未形成系统。数据与科学研究范式的密切关系,完全有理由从哲学角度进行系统反思。我们尝试构建数据哲学的框架。

四 构建数据哲学的基本框架

数据哲学主要研究数据的一般规律、获取和使用数据的基本方法及其发展中的哲学问题、数据在科学技术和社会发展中的作用等内容。由于数据在科学技术活动和其他社会活动中的重要地位已经日益凸显,将数据作为一个单独对象考察和研究无论对科技发展还是对社会发展都具有重要的作用。数据哲学研究将涉及到科学技术史、自然哲学、科学哲学、技术哲学、思维科学、科技社会学、科技方法学、科技伦理学、技术经济学等多个学科。在此我们以与科研活动有关的科学数据为例阐述构建数据哲学的理由。

首先,将数据纳入到科学哲学研究范畴之中的时机已经成熟。科学哲学的研究内容自诞生以来得到不断的充实和发展,在每一次科学哲学内容的变革中,科研活动所取得的成就起到了决定性作用。早期的传统科学哲学将科学理论的产生和发展作为自己的研究任务,探讨的是科学理论的形成和对科学现象的解释,而忽略了观察事实的形成过程。20世纪的物理学革命,尤其是量子力学的诞生打破了人们的牛顿力学体系能对自然进行确定性研究的传统观念,引发了科学哲学领域的实在论与反实在论的激烈争辩,拓展了科学哲学的研究内容,与理论、观察相关的其他科学研究要素逐渐成为科学哲学的研究对象,比如测量问题的测量哲学成为科学哲学探讨实在论的重要手段,拓展了科学哲学研究的空间。

在传统科学哲学中,数据仅仅被看成是科学研究的自然结果,是研究对象特征的表征,哲学家关注事实、知识和对知识的理解,很少将数据作为研究的对象,以至于将观察事实成为数据的代名词,掩盖了数据的作用。近代科学诞生的重要标志就是强调观察测量的重要性,将数据纳入到研究体系之中,随

着科学的发展,数字化时代中数据的地位和作用也发生了相应的变化。

近年来,贯穿于科学研究过程的数据逐渐被哲学家们所关注,如对数据与现象的关系、理论与数据的关系等系列问题的探讨。然而,对于数据的系统的哲学研究至今尚未形成。事实上,无论是未经处理的原始数据,还是最后形成科学知识一部分的数据,数据始终是科学研究中必不可少的重要因素。无论是理论的形成,还是对理论正确性的判断,直至理论的变更,都与数据有密切关系。数据更是与观察、观测紧密相连。在当代科学和社会活动中,数据正在产生着巨大影响。因此,我们将数据纳入到科学哲学的研究领域,系统地研究科学家所提供的数据解释和形成理论的哲学基础,具有十分重要的现实意义,而且为丰富科学哲学的研究提供了一条新路径。

其次,与数据密切相关的数和量都已经有了相应的哲学研究领域,为构建数据哲学提供了相应的理论基础。以数为研究对象的学科是数学,相关数的哲学研究属于数学哲学之列,也形成了相应的学派。在认识数的过程中,历史上有几个重要流派^[34]:柏拉图主义——数存在于理念世界;唯名论观点——数是纸上的符号或头脑中特定的概念;康德——数是思维创造的抽象实体;以及现代数学的三大流派——逻辑主义、直觉主义、形式主义。与数形成鲜明对比的是量,它以数为基础,对可观察和可测量的事物进行量化,数可以是抽象的,而量是具体的,量是对物质特性的具体描述。对量的研究是定量研究的必然结果,近代科学的最大特点之一就是量化研究以提高研究的精确度,因此对量的研究遍及各个学科领域。量对科学研究如此重要,必然引发一系列的哲学问题;量子力学的诞生,打破了人们树立的牛顿力学体系能对自然进行确定性研究的传统观念,尤其是测量结果的实在性问题引发了测量哲学的诞生。

与数学哲学和测量哲学相对应,构建以数据为研究对象的数据哲学,可以使其成为科技哲学的一个研究方向。科学数据的高度纯粹的客观性是科学研究追求的目标和期望,科学数据的准确性和可靠性是一切科学研究和社会发展的基础。然而,由于科学研究自身的不确定性、科学共同体认知范围的局限性、以及社会环境对数据的客观性和准确性都会产生巨大的影响。从认识论的角度解读数据与现象、理论、观察的关系,从方法论角度解释数据在科学研究中的作用和地位,以及对数据的本体论认识和价值论认识都将成为数据哲学研究的基本内容。

最后,在科学哲学研究内容不断丰富的同时,其研究方法和手段也向多元化方向发展,后逻辑实证主义不仅将科学哲学的研究重点转向为研究科学研究中的实践操作过程及其客观性问题,而且还将科学研究置于社会、文化等环境中,催生和发展了以科学社会学和科学文化学为代表的众多学科,他们的主要研究方法是通过跟踪科学实验或者对科学史进行历史考察,探讨在一定的社会环境中,科学研究过程中理论与观察之间的关系。

科学技术史、科学社会学、科技伦理学和科学文化学等学科的研究范式为我们研究数据的哲学研究拓展了思路。除利用跟踪科学家和使用科学史的研究方法,还可以借助管理学、社会学等诸多方法,解决近年来科学数据在其获取、解读、共享、伦理和管理方面急需解决的大量问题。

总之,数据哲学为我们提供了极其广泛的研究内容。作为科学哲学的新分支,数据科学哲学将以数据为基本范畴,聚焦于数据的内涵与性质、产生与挖掘、利用与共享等问题,在汲取数学哲学和测量哲学的相关成果基础上,不仅着眼于探讨其本体论、认识论、方法论和价值论方面的问题,还可辅之以科学史、伦理学、社会学和文化学视野的理论探讨。

注 释

[1]DC 研究报告 《2011 年全球数据总量》1.8ZB,2012 年 8 月 6 日。<http://it.zbinfo.net/tech/383219/602816813154.shtml>

[2][13][30]M. Joshi and S. Krag, "Issues in Data Management", *Science and Engineering Ethics*, 2010, 16 (4), pp. 743 - 748. (Special Issue: Responsible Data Management).

- [3]《美国政府将投入 2 亿美元加强“大数据”研究》，2012 年 8 月 6 日。<http://news.csdn.net/a/20120330/313722.html>
- [4]<http://mymemory.translated.net/>
- [5]中国社会科学院语言研究所词典编辑室《现代汉语词典》，商务印书馆，2002，第 1176 页。
- [6]<http://en.wikipedia.org/wiki/Data>，<http://zh.wikipedia.org/wiki/%E6%95%B0%E6%8D%AE>。
- [7]Online etymology Dictionary http://www.etymonline.com/index.php?allowed_in_frame=0&search=data&searchmode=none。
- [8]Francis Galton, “Biometry”, *Biometrika*, 1901, 1 (1), pp. 7–10.
- [9]Jerome M. Clubb, Erik W. Austin, Carolyn L. Geda, and Michael W. Traugott, “Sharing Research Data in the Social Sciences”, Stephen E. Fienberg, Margaret E. Martin and Miron L. Straf eds., *Committee on National Statistics, National Research Council, Sharing Research Data*, National Academy Press, 1985, p. 42.
- [10]Stephen E. Fienberg, Margaret E. Martin and Miron L. Straf eds., *Committee on National Statistics, National Research Council, Sharing Research Data*, National Academy Press Washington, 1985, pp. 7–8.
- [11]Andrew J. Severin, “Dealing with Data, Training New Scientists”, *Science*, Vol, 331 25 March, 2011, P. 1516
- [12]朱扬勇、熊贻《数据学》，复旦大学出版社，2009。
- [13]J. Bogen, J. Woodward, “Saving the Phenomena”, *Philosophical Review*, 1988, 97 (3), pp. 303–352.
- [14]James W. McAllister, “Phenomena and Patterns in Data Sets”, *Erkenntnis*, 1997, 47, pp. 217–228.
- [15]Bruce Glymour, “Data and Phenomena: A Distinction Reconsidered”, *Erkenntnis*, 2000, 52, pp. 29–37.
- [16]J. Apel, “On the Meaning and the Epistemological Relevance of the Notion of a Scientific Phenomenon”, *Synthese*, 2011, 182 (1), pp. 23–38.
- [17]S. Schindler, “Bogen and Woodward’s Data-phenomena Distinction, Forms of Theory-ladenness, and the Reliability of Data”, *Synthese*, 2011, 182 (1), pp. 39–55.
- [18]U. Feest, “What Exactly is Stabilized when Phenomena are Stabilized?”, *Synthese*, 2011, 182 (1), pp. 57–71.
- [19]J. Worrall, “Underdetermination, Realism and Empirical Equivalence”, *Synthese*, 2011, 182 (2), pp. 157–172.
- [20]J. Sprenger, “Science Without (Parametric) Models: The Case of Bootstrap Resampling”, *Synthese*, 2011, 182 (1), pp. 65–76.
- [21]J. Bogen, “‘Saving the Phenomena’and Saving the Phenomena”, *Synthese*, 2011, 182 (1), pp. 7–22.
- [22]P. Machamer, “Phenomena, Data and Theories: A Special Issue of Synthese”, *Synthese*, 2011, 182 (1), pp. 1–5.
- [23]B. Lowe, T. Muller, “Data and Phenomena in Conceptual Modeling”, *Synthese*, 2011, 182 (1), pp. 131–148.
- [24]B. Falkenburg, “What are the Phenomena of Physics?”, *Synthese*, 2011, 182 (1), pp. 149–163.
- [25]K. Brading, “Autonomous Patterns and Scientific Realism”, *Philosophy of Science*, 2010, 77 (5), pp. 827–839.
- [26]LB. Makeeva, “Scientific Realism, Truth, and the Underdetermination of Theories by Empirical Data”, *Russ. Stud. Philos*, 2010, 49 (3), pp. 58–71.
- [27]西斯蒙多《科学技术学导论》，许为民等译，上海世纪出版集团，2005。
- [28]Joe Giffels, “Sharing Data is a Shared responsibility”, *Science and Engineering Ethics*, 2010, 16 (4), pp. 800–803.
- [29]D. Crome, “Avoiding Twisted Pixels: Ethical Guidelines for the Appropriate Use and Manipulation of Scientific Digital Images”, *Science and Engineering Ethics*, 2010, 16 (4), pp. 639–668.
- [30]C. Andrew, “Rescue of Old Data Offers Lesson for Particle Physicists”, *Science*, 11 February, 2011, 331, p. 694.
- [31]R. Sarah, “Is There an Astronomer in the House”, *Science*, 11 February, 2011, 331, p. 696.
- [32]T. Jonathan, “Overpeck, Gerald A. Meehl, Sandrine Bony, and David R. Easterling, Climate Data Challenges in the 21st Century”, *Science*, 11 February, 2011, 331, p. 700.
- [33]<https://www.citiprogram.org/>
- [34]张景中、彭鑫成《数学哲学》，北京师范大学出版社，2010。

(责任编辑 徐 兰)