

# 大数据中的因果关系及其哲学内涵

王 天 思

摘 要：在大数据中，数据化使因果关系量化为变量之间的关系，在获得关系强度和正负性质的同时，丧失了原有的必然性和方向性。大数据的相关关系，进一步展开了因果概念的重新刻画：因果关系是对因素相互作用过程与其效应之间关联的描述；而相关关系所描述的则是因果派生关系。作为因果派生关系，相关关系根植于因果性；作为未进入相互作用过程凝固为因果关系的因素关系，相关关系提供了由因素创构结果的广阔空间，这正是数据物化的因果性根据；而作为因素分析，相关定量分析的因果派生依据则构成大数据分析的因果基础。大数据中因果关系的厘清，晓示了其深层哲学内涵。因素关系的未来空间凸显创构认识论，因果派生关系的全数据定量分析呈现量的整体把握，而因果关系从描述到创构则彰显哲学以满足人的需要为最终目的。

关键词：大数据 相关关系 因果关系 数据物化

作者王天思，哲学博士，上海大学社会科学学院哲学系教授（上海 200444）。

“大数据”（big data）概念最早出现于《科学》杂志刊载的《大数据的处理程序》（1998）一文，此后《自然》杂志于2008年9月出版“大数据”专刊，短短十数年，“大数据”浪潮以人们难以想象的速度和让人始料未及的方式袭来。大数据一方面对传统的因果关系理论提出了挑战，另一方面又给因果概念的进一步展开提供了更高层次的整体观照。作为一种信息存在方式，大数据中因果关系和相关关系问题的错综交织，成为当前需要面对和思考的哲学课题。

## 一、相关关系的因果性根基

关于大数据的看法，最引人注目的观点无疑是强调要注重相关关系而忽略因果关系，甚至某种程度上认为相关关系可以取代因果关系。

### （一）因果性和相关性的关系问题

当人们正掂量“依赖模型的实在论”<sup>①</sup> 还有多少实在感时，有人基于大数据，从“模型都是错的，只是有些有用”，<sup>②</sup> 推出了理论终结的结论。30 年前基于统计模型得到的模型观，今天在大数据时代被赋予了新的含义：模型的过时，意味着“理论的终结”。2008 年，美国《连线》杂志主编克里斯·安德森（Chris Anderson）以《理论的终结》为题发表论文，认为“有了足够的数据，数字会自己说话”，因而“相关关系足够了”。“相关关系取代因果关系，而科学甚至可以没有连贯一致的模型、统一的理论，或者实际上可以完全没有任何机制性解释而发展”。<sup>③</sup> 一些被看作是从大数据推出，因而影响广泛的观点认为，“大数据时代最惊心动魄的挑战，就是社会将需要从对因果关系的某种痴迷中蜕出，而代之以简单的相关关系。在大数据基础上建立起来的模型没有关于‘为什么’的理解，只有‘是什么’。大数据是关于‘是什么’而不是‘为什么’的”。这种观点甚至断言，“确认因果机制是一种自诩得意的幻相，大数据推翻了这个假象。我们又一次陷入一个历史绝境：在这里‘上帝死了’。”<sup>④</sup> 这些现在比较流行的观点，既涉及大数据时代的重要观念转变，同时也与因果关系的认识迷雾密切相关，而且往往两方面相互交织。迷雾出自因果关系的传统理解，而重要识见则源于信息时代认识的深化。

因果关系是人类理性行为与活动的基本依据，人类理性本身不可能否定因果关系，但大数据所凸显的相关关系，的确从实践层面实质性地推进了对传统因果概念的深入反思。随着思考的深入，重重迷雾中的问题所在日渐明确：“大数据中一个耳熟能详的说法是：大数据长于分析相关关系，而非因果关系。但这可能是一个伪命题。如何从相关关系中推断出因果关系，才是大数据真正问题所在。”<sup>⑤</sup> 目前亟待探索和解决的，正是有关大数据的这一问题。而这一问题的解决，还在于因果性和相关性问题的澄清。

① 史蒂芬·霍金、列纳德·蒙洛迪诺：《大设计》，吴忠超译，长沙：湖南科学技术出版社，2011 年，第 4 页。

② George E. P. Box, “Robustness in the Strategy of Scientific Model Building,” in R. L. Launer and G. N. Wilkinson, eds., *Robustness in Statistics*, New York: Academic Press, 1979, p. 202.

③ Chris Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” *Wired*, June 2008.

④ 参见 Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution that Will Transform How We Live, Work and Think*, New York: Houghton Mifflin Harcourt, 2013, pp. 7, 12, 17.

⑤ 姜奇平：《因果推断与大数据》，《互联网周刊》2014 年第 18 期。

对于因果关系来说，物的数据化过程意味着特殊的遭遇。因果关系的这种遭际，无疑是大数据的发展带来的，但问题的根源却还在因果观念本身的发展困境。由于“本原”的辉映，古希腊哲人认为智慧是关于原因的科学，因而关于因果观念，最先关注的主要是原因而不是因果之间的内在关联。无论亚里士多德的“四因说”还是东方的“因明学”，都是主要关注原因的典型形态。休谟第一个认识到，要给原因下一个完备的定义，必须呈现其与结果的联系，但他发现要做到这一点极为困难：“关于这一联系的观念，当我们努力去构想它时，甚至连我们究竟想要知道它的什么内容，都没有一个哪怕是模糊的意识。”<sup>①</sup> 因此在其经验研究进路中，休谟只能建立起原因和结果的时序关联。康德试图通过先验进路探索这一问题，他把因果观念从“自然的”转向“理知的”，<sup>②</sup> 但因果间内在关联的先验解决，只是把因果问题推给了先验问题。休谟因果观念研究的经验进路，在经验科学中导向了因果关系的概率研究；而康德的先验进路则通过无所不在的因果假设，晓示着自己不可或缺的逻辑地位。由于因果观念的康德理解封存于先验，在穆勒之后因果关系研究的发展中，出现了因果关系科学研究和哲学研究的明显分化。

从穆勒开始，作为归纳的基础假设，因果观念逐渐纳入逻辑研究，后经罗素、赖辛巴赫和卡尔纳普等，将因果关系的研究由必然性退离至或然性。由于作为客观世界本身的特性，或然性规定构成不可克服的内在逻辑矛盾，这种研究进路到波普那儿走向主体“大胆的猜测”。与逻辑学研究相应，因果观念研究的另一致思方向则是统计物理学、统计学、计量经济学、生物统计学和心理测验学等学科中的量化研究进路。在因果关系的量化研究进路中，高尔顿的线性回归模型意味着从严格意义上的“物理模型”转向“统计模型”；皮尔逊的“相关系数”、凯恩斯的“归纳相关”及基于预测的“格兰杰因果关系”，则使相关关系研究在经验科学中占主导地位。而当这种进路深入到基础科学特别是量子理论时，因果性研究则遇到了量化把握所无力应对的根本挑战和困难。如果说因果观念的定性研究困在了经验和先验之间，那么，因果关系的定量研究则惑在因果关系和相关关系之间。

因果关系研究的量化，必定走向主要把握相关关系。在大数据中，物数据化就意味着因果关系的数据化。在因果关系的数据化过程中，一方面由于量化获得了量的关系强度和正负性质；另一方面由于量化丧失了原有的必然性和方向性。因此，量化后的因果关系蜕变为变量之间的相关关系，大数据中物数据化的过程，同时是因果关系量化为相关关系的过程。正因为如此，物数据化的结果，一方面有利于因果关系的定量把握；另一方面又不利于因果观念的传统定性理解。也正因为如此，

① David Hume, *An Enquiry Concerning Human Understanding*, New York: Oxford University Press Inc., 1999, p. 146.

② 康德：《纯粹理性批判》，邓晓芒译，北京：人民出版社，2004 年，第 440 页。

数据化基础上的因果推断就变得越来越重要。而因果推断重要性日增，本身就表明，不管对因果性持什么态度，客观上都不可能否定因果关系。

因果关系不可能被否定，而大数据时代又似乎要以相关关系取代因果关系，那么，问题只能在传统因果观念中。在哲学上，传统因果概念备受质疑；在现代科学尤其在量子理论中，已经导向因果概念的重新刻画；而大数据所涉及的相关性和因果性之间的关系问题，则使重新刻画的因果概念进一步展开并系统生效。

## （二）因果关系是对因素相互作用过程与其效应之间联系的描述

20 世纪 30—50 年代，对于因果性问题来说，是灾难性的 30 年。因果论遭遇的这种灾难，无疑主要是由量子理论带来的，但量子力学本身并不否定因果性，而只是要求对传统因果概念进行改造。重新刻画因果概念，既是现代科学发展的要求，也是克服传统因果概念内在逻辑矛盾的需要。关于因果概念研究的这种困境，集中体现于两个关键问题：一是因果联系是必然的还是或然的；二是因果联系是前后相继的还是同时的。

因果联系的必然性和或然性问题，涉及因果规定的合理性。在休谟的观念中，“绝对不可能”在定义“原因”时，可以不在其涵义中包含原因与其结果之间所具有的那种“必然性联系”。<sup>①</sup> 在康德作为先天形式的因果概念中也是如此：“原因的概念，它陈述的是在某种前提条件下一个结果的必然性。”的确，在哲学中，我们必须认为“一个原因的概念就带有必然性的特征”，<sup>②</sup> 否则我们就不能理解事物过程的确定性一面。而在科学领域情况则很不相同，不仅统计学和计量经济学，量子力学也必定使传统因果观不得不接受原因和结果之间具有或然联系的观点。然而，如果原因和结果之间的联系是或然的，那么世界就毫无疑问是变幻莫测的，事物过程就不会有确定性环节，从而也就很难为人类理性所把握。但是，如果我们不接受原因和结果之间或然联系的观点，传统因果概念就不能容纳任何性质的统计理论，更不能描述量子现象。

而因果时序问题，则涉及传统因果概念的内在逻辑矛盾。关于因果时序，向来存在“前因后果”说与“因果同时”说之争。在量子理论创立之后，又出现了危及因果性本身的“因果时序颠倒”说。因果相继的观念几乎与因果观念同时并存，早在休谟关于因果观念的考察中，就把“前因后果”视为当然。在康德那里，作为先天形式的因果观念就先验地具有前后相继的性质。而在因果关系的量化研究中，由于因果关系量化为变量之间的关系，“自变量”和“因变量”的概念本身就意味着以前后相继为前提性预设。在传统因果概念中，前因后果说和因果同时说都具有其合

<sup>①</sup> David Hume, *An Enquiry Concerning Human Understanding*, p.159.

<sup>②</sup> 康德：《纯粹理性批判》，第 111、123 页。

理性，但二者又构成传统因果概念的内在逻辑矛盾。“前因后果”说认为，原因的“始发性”和结果的“后继性”使因果联系具有时间上的先后顺序性。但“因果同时”说却认为，“前因后果”说在理论上具有不可克服的困难。因为它必然隐含着“前因”阶段存在无果之因的结论。这既不合乎辩证法，也与因果关系的基本规定相矛盾。要避免这些内在逻辑矛盾，就必须重新刻画因果概念。只有重新刻画因果概念，才能在确保因果关系必然性联系的同时，既使因果关系具有时间上的持续性，又使事物过程的某些环节具有或然性的一面。

传统因果概念的根本缺陷，就在于把原因等同于某种原子式的存在，而没有把它看作对事物相互作用过程的描述。把因果概念理解为原子式的现象产生另一个原子式的现象，源自对因果概念的日常或经典力学理解。只有把因果概念的规定落实到事物的相互作用上，才可能有与当代科学发展相适应的因果概念规定。如果我们把参与相互作用过程、构成一定原因的事物称为“因素”，即构成原因的要素，那么，传统因果观某种类似原子而不是一个相互作用过程的“原因”，实际上只是“因素相互作用过程”中的因素，而不是作为因素相互作用过程意义上的原因，不过常常是主要因素罢了。因素关系之所以不同于一般意义上的事物关系，关键就在于它与特定的结果联系在一起，而这种结果不仅可以是既存的，也可以是潜在的。而原因和因素的根本区别则在于：原因由因素构成，但原因不仅包括因素本身，更重要的是因素的相互作用过程。正是在这个意义上，原因是对因素相互作用过程的描述。

把原因看作对因素相互作用过程的描述，就为进一步刻画“结果”概念创造了条件。相互作用都伴随着相应的效应；原因作为因素的相互作用过程，每一瞬间也同时伴随着这种相互作用所产生的效应。作为对因素相互作用产物的描述，这种效应自然是最为直接的结果，即因素相互作用过程的“瞬时结果”。这种瞬时结果与我们通常所谓“结果”并不完全是一回事。作为日常概念的“结果”，往往指的是这些效应的痕迹累积而成的宏观事物或现象，甚至指的是整个因素相互作用过程的产物。作为整个因素相互作用过程的结果，自然不是指某种瞬间状态，而是一种轨迹式的连续状态或事物系统。

效应和结果是两个既密切相关又不完全相等同的概念。在英文中，效应和结果用同一个词“effect”表达，这增加了对词义理解的复杂性。事实上，作为严格刻画的因果概念，结果只能是作为瞬时结果的效应。某一原因消失，即某些因素的相互作用停止，效应便同时终止，但其痕迹却可以依然存在。因此结果可以定义为：因素相互作用的效应；更确切地说是因素相互作用效应的描述。因素和结果间的联系与原因和结果间的联系，描述的是客观过程中处于两个不同发展阶段的事物联系形式。其中，因果关系描述的是因素相互作用过程与其效应之间的联系。

通过深化因果观念的理解，重新刻画的因果概念解决了传统因果观存在的上述两大问题。如果说，关于因果联系的逻辑方面的理论混乱，是由于把因素看作原因，

那么，关于因果联系时序方面的上述纷争，则主要是由于把因素相互作用所产生的效应，与这些效应的痕迹累积而成的结果混为一谈。当我们把原因看作对因素相互作用过程的描述，从而将原因和因素区分开来研究，就会看到，原因和结果是因素相互作用过程的两个不同方面，因果联系的必然性无可置疑，与原因和结果间的关系不同，因素（包括潜在的）与结果间的联系从根本上说是或然的，事物过程必然性和或然性的对立在因素相互作用过程中获得了统一的理解。因果描述所涉及的联系的过程性，实际上是因素相互作用的过程性；而原因所描述的能动性，则源于因素活动的始发性。<sup>①</sup> 在原因是因素相互作用过程的描述中，可以看到“前因后果”的观念出于两方面误解：一是把主要因素当作原因，因而可以与其参与相互作用构成的结果有时间先后关系；二是把因素相互作用的效应累积的产物当作结果，这也使结果可以在时间上后于原因。重新刻画的因果概念，不仅能使传统因果观的上述问题得以澄清，而且意味着因果关系和相关关系之间一个同样清晰的关系刻画。

### （三）相关关系是对因果派生关系的描述

从定性研究的角度看，相关关系是一个宽泛的概念，因为从直接相关到间接相关，呈现的是一个无限蔓延的分布。世界上任何两个事物都处于相互联系之中，都具有相关性。因而从广义上说，相关关系可以把任何事物视为相互联系，因为关系本身就是对相关性的描述，这里显然存在有待澄清的理论问题。

关于相关关系和因果关系，自 17 世纪以来就有一个影响广泛的观念：“相关关系并不意味着因果关系。”<sup>②</sup> 20 世纪后，出现了将因果关系归结为相关关系的努力，认为“因果关系不是别的，只是一类相关关系”。<sup>③</sup> 将因果关系归结为相关关系，显然将因果关系的理解泛化了，结果是因果关系弥散于相关关系。但是，这种观点是第一次内在沟通因果关系和相关关系的思考。这种思考，后来在赖辛巴赫那儿得到另一方向的强化，他认为“不存在没有因果关系的相关关系”，<sup>④</sup> 这则是出于对相关关系因果关联的确信。这种理解更接近二者关系的实质，但由于相关关系过于宽泛，因而不仅涉及相关关系的规定问题，而且涉及相应的哲学基本问题。而关于这一问题，统计学和计量经济学等学科中的认识更为具体：“相关系数度量的是相互联系，

① 参见王天思：《哲学描述论引论》，上海：上海人民出版社，2009 年，第 331—333 页。

② Robert R. Pagano, *Understanding Statistics in the Behavioral Sciences*, 10<sup>th</sup> ed., Wadsworth: Cengage Learning, 2013, p.144.

③ 参见 Judea Pearl, *Causality: Models, Reasoning and Inference*, 2<sup>nd</sup> ed., Cambridge: Cambridge University Press, 2009, p.176.

④ Hans Reichenbach, *The Direction of Time*, Berkeley: University of California Press, 1956, p.44.

但相互关系并不等于因果关系。”<sup>①</sup>“相关关系并不能证明因果关系。”<sup>②</sup>所有这些观点都聚焦于一个环节：“建立有意义的相关关系是一回事，而实现从相关关系到因果属性的飞跃则是另一回事。”<sup>③</sup>当我们重新刻画因果概念，就可以看到这一问题的关键所在。

重新刻画的因果概念为相关关系的深入理解提供了理论根据，为这些问题的澄清创造了条件。在重新刻画的因果概念中，因素和原因的区分打开了通向相关关系的窗口。当因素未进入相互作用过程时，我们可以看到一种特殊的关系，那就是与潜在结果相联系的因素关系。这种关系不是因果关系，但相对于潜在结果，却构成一种与因果关系密切相连的相关关系。这种相关关系的生成，表现为一种因果派生关系。这不仅表明，因素之间的相关关系并不是一种与因果关系无关、完全不同的另一种关系，而且可以看到，因果结构可以派生出更多相关关系。因果派生关系除了因素之间，还有结果之间以及因素和结果之间的关系等，这就可以看到由不同的机制可以派生出各种不同类型的相关关系。

作为因素相互作用过程与其结果之间的关系，因果关系是由相互作用过程凝固了的因素关系，而在进入相互作用之前和因素相互作用过程之中，因素之间、结果之间以及因素和结果之间则构成了各种类型有特殊意义的关系，这些都是相关关系。正是由此，可以得到关于因果关系和相关关系的清晰理解：因果关系是为因素相互作用所确定的关系，由于因果关系事实上是一个过程的两个方面，所以不构成相关关系。而相关关系则是因果关系的派生关系，包括因素之间、结果之间以及因素和结果之间的关系等。虽然因果关系和相关关系有重要区别，但从重新刻画的因果概念，可以清楚地看到相关关系的因果性根基。作为因果派生关系，相关关系植根于因果关系。

正是在这个意义上，大数据不仅不是因果性观念的沦陷之地，反倒是进一步充分展开重新刻画的因果概念的最好场所。正是大数据带来了物数据化和数据物化的对称发展，而数据物化既是实践又是认识，它意味着一种新的因果性根据。

## 二、数据物化的因果性根据

作为人类信息文明的基础，大数据时代的发展过程由两个基本方面构成：一是

① David Freedman 等：《统计学（第二版）》，魏宗舒等译，北京：中国统计出版社，1997年，第169页。

② Jessica M. Utts and Robert F. Heckard, *Mind on Statistics*, 4<sup>th</sup> ed., New York: Duxbury Press, 2011, p. 94.

③ David Bollier, *The Promise and Peril of Big Data*, Washington, DC: The Aspen Institute, 2010, p. 16.

物数据化；二是数据物化。数据物化意味着从关注因果关系到关注因素关系等相关关系，而从对因果关系的关注到对相关关系的关注，甚至出现以相关关系取代因果关系的倾向，则事实上是在还没有理清因果关系和相关关系之间关联的基础上，对二者的割裂理解，因而只是意味着从主要注重描述到同时注重创构。

#### （一）大数据相关关系提供了由因素创构结果的广阔空间

由于涉及大数据，有一个重要环节可以大大推进我们关于因果关系的认识。如果把原因——也就是因素相互作用过程返演，就会发现一个很有意义的情景：在进入相互作用过程之前，因素还没有构成原因，这时候表现为具有某种结果指向的因素和因素之间的关系——大数据所最具价值的相关关系之一。正是这种相关关系，给创构特定的结果留下了广阔的实践操作空间。

作为与既存结果相联系的因素关系，相关关系是我们解开既存结果——事物现象的密码；而作为与潜在结果相联系的因素关系，相关关系则为我们创构所想要的结果提供了广阔空间。由于没有进入相互作用过程生成既定的结果，因素关系为人类活动提供了空间，使人们能由此对因素相互作用可能生成的结果进行选择，通过引导这种因素关系，使它们以特定的方式进入特定的相互作用，从而得到所想得到的特定结果。对与既存结果相关联的因素相互作用的描述是过去时态，而对与潜在结果相关联的因素相互作用过程的描述则是将来时态。从因素构成就可以看到，因素关系或相关关系之所以特别重要，大数据之所以重要，关键之点就在于：人们对因果关系的把握从既存结果推展到潜在结果，从过去时推展到将来时，从而能够根据因素关系进行结果创构。

在因素构成原因的过程中，相互作用过程一返演，我们看到的就是因素，这时候因素所构成的是相关关系，而相关关系给我们提供了一个前提、一个基础，我们可以利用这些因素，通过组合让它们相互作用，得到我们想要的结果。这是一个可以让人类拥有无限可能性的空间，由此足见因素进入相互作用凝固相互关系前后的重要不同。大数据给我们提供了一个重要平台，一个信息文明时代人类创造活动的新舞台。正是这个平台，蕴含着大数据时代数据物化的因果性根据。

作为物数据化的结果，大数据是数据物化的前提。物的数据化只是信息文明时代的前奏，真正的信息文明时代是数据的物化，那就是刚刚出现在我们生活中的“3D打印”和“谷歌眼镜”等所谓“新硬件”。它所开启的是一个创构的时代，所谓“新硬件时代”，事实上是创构时代的序幕。这意味着人类文明从一个以描述活动为主的时代，进入到一个以创构活动为主的时代。尽管有史以来，描述和创构总是相伴而生的基本活动，但信息时代特别是大数据凸显了一种与描述活动不同的创构活动，它不仅意味着数据物化的因果性根据，而且意味着因果性认识的一个重要发展。



数据物化是一个与物数据化相反的过程,在这一过程中,结果不是既定的,因而既不是由果溯因,也不是一般意义上的由因(素)索果,而是由因(素)构果。这个“果”也不是既存的,而是根据人们的需要和大数据相关关系所反映的可能性,通过以一定的方式使相关因素进入相互作用过程,所获得的期望中的结果。而在特定潜在结果的观照中,则会生成新的潜在的因果关系。这种潜在因果关系所构成的因素体系,在还没有进入相互作用以前,只具有因素和潜在结果之间的关系。只有让相关因素以一定的方式进入相互作用,生成预期中的结果后,因素相互作用过程才构成现实的原因,从而生成现实的结果,形成作为潜在物创构过程的因果关系。这种因果关系不同于既定物生成的因果关系。正是由此可以看到描述和创构两个不同的空间和过程,看到两种不同的认识形成的必然性和合理性。

## (二) 创构活动的因果性基础

几乎从一开始,人类就有两类基本的活动:探索既存事物和创设以往不存在的全新事物。这两种不同的活动需要两种不同的认识:对既存事物或世界存在规律的认识和对创设新事物规律的认识。前者主要是描述性质的,是一种描述活动;而后者则主要是创构性质的,是一种创构活动。

描述是对既存对象的符号、语词和模型等的摹写;创构是从未存在的感性对象的创设,它是基于人们的需要,根据所设立的潜在结果,得到与之相联系的因素体系,并确定和控制这些因素的相互作用,获得所想得到的结果的过程。之所以说“感性对象”,既为区别于观念及其体系,也为包括由原子构成的新硬件等实物和由比特(bit)构成的虚拟物,即信息对象。而数字虚拟对象更典型地诉诸人的感觉。创构不仅是一个在物数据化基础上的数据物化过程,而且是数据通过虚拟对象化的过程。

由于描述更多是尽可能真实、精确地反映既存对象,因此描述活动的主要特征是用思想重建既存对象,更以真实性为衡量标准,主要基于现实的因果关系追问“为什么”。由于创构更多是根据人的需要及其发展创设满足和开发人的需要的感性对象,典型的是新硬件和虚拟实在,因此创构活动的主要特征是创设从未存在的可感对象,更以满足主观需要为主要目的,以使用的有效性为主要衡量标准,主要基于面向可能结果的因素关系,创设新的“是什么”。

数字技术的发展,使人类的创构活动进入了一个新阶段。而在大数据基础上,人类的实践和认识活动主要是创构性质的,描述和创构是信息文明时代一对重要的基本概念。<sup>①</sup>从另一方面看,也就是大数据为创构活动提供了新的基础。正是大数

<sup>①</sup> 参见王天思:《描述和创构——关于数字时代哲学走向的两点思考》,《江西社会科学》2004年第1期。

据时代，与描述性认识不同的创构性认识空前凸显，在人们的意识中，创构意义上的因果性与描述意义上的因果性出现对立和冲突，从而构成了大数据浪潮对单纯描述意义上的传统因果观的巨大冲击。

由于是凝固了的相关关系，因果关系对于创构来说没有操作价值，只有原理参照价值或作为模拟对象的价值。而作为没有凝固的可能因果关系，相关关系虽然不是既定事实，却具有重要实际操作意义。因而描述性认识主要基于因果关系，更属于因素已经通过相互作用凝固为结果后的由果溯因；而创构性认识则主要基于相关关系，更属于因素还没有通过相互作用凝固为结果而与可能结果相联系的由因（素）构果。

正是在这里，深入涉及数据物化的因果根据；也正是由于对创构的因果根据认识不足，才造成了相关关系和因果关系理解上的两方面问题。一方面，在物能相互作用中，我们一般总倾向于把主要因素看作就是原因，而且这样在日常生活中使用起来非常方便，但在大数据时代，由于因素众多且极其复杂，我们不再能把主要因素看作就是原因本身了。另一方面，关于因果关系和相关关系的割裂理解，正是只顾及创构活动而没有涉及其因果根据的结果。由于创构总是与对外部世界的描述具有某种根本的相关性，因此只顾及创构而不管其本体论基础是相对短视的。只要联系到创构产物跟人的关系，就能感觉到其中的短视所在。这里所涉及的远不只是对因果关系的消解倾向，更涉及从“为什么”到“是什么”的创构，涉及创构的因果根据。

认为大数据表明“是什么”，而不是“为什么”，也是由于没有区分描述和创构，更未辩明二者之间的关联，尤其是没有涉及描述和创构的共同因果根据。这里，也涉及不同层次的两方面问题。一方面，描述必须知道“为什么”，而创构却未必。只要“是什么”，不要“为什么”的观点，正是由此而来。而这种误解，恰恰又因为未深究“是什么”和“为什么”之间的深层次区别和联系。另一方面，由于不清楚二者的区别和联系，就不可能意识到更深层次的重要问题：在创构性认识层次，“是什么”有时候不仅可以，而且越是新的创构越必须建立在“为什么”的基础上，因而还可能包含比通常意义上的“为什么”更深刻的内容。

### （三）建立在“为什么”基础之上的“是什么”

的确，亚马逊创新的推荐系统只是梳理出了有价值的相关关系，而不知道其背后的原因，但由此得出“知道是什么就足够了，不用知道为什么”<sup>①</sup>的结论，则只是就亚马逊卖出商品而言，只是因为局限在一个有限的目标。在一个有限的活动内，可以不管超出这一目标的因果关系。但问题在于，即使纯粹在大数据基础上，人类活动也仍然是一个具有不同层次的整体。一方面，不需要知道“为什么”并

<sup>①</sup> Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution that Will Transform How We Live, Work and Think*, p.51.

并不意味着不存在因果性，我们要的是预想的结果，因而重要的是因素，重要的是什么样的因素通过什么样的相互作用可以得到什么样的结果。而另一方面，在创构活动中，创构的产物不是既存的东西，而是新的“是什么”。这种新的“是什么”，无疑必须建立在“为什么”的基础之上。这方面，基于大数据和基于小数据没有原则区别。

正是在创构的意义上，大数据时代和小数据时代具有同样的性质。不仅在小数据时代，大数据时代也需要假设，需要假设的实验验证，只是在数据层面可以忽略假设，而让我们所需要的结果呈现，就像在创造发明时可以相对忽略本体论。大数据的确可以减少很多假设，因为在数据覆盖范围内，就数据所能达到的认识而言，不需要太多假设弥补逻辑缺环，但那只是就大数据的实际操作层面而言。而操作要可靠，要更有效地操作，还必须有更深层次的根据，就像在量子力学数学形式体系层面，我们可以不借助物理解释，但量子理论本身却必须有对其物理意义的理解。

在量子力学中，人们可以运用数学形式体系，通过哪怕是很小几率的确定性，就能得到预期中或想要的结果，但这并不意味着从此人们可以没有关于量子对象的物理解释，或者说有了量子力学的数学形式体系就足够了。即使像在量子力学那样的基础科学中，人们都不能满足于数学形式体系的有效性甚至精确性，在大数据时代，人们更不可能完全满足于表面的相关性，除非局限在某个有限目的或需要范围，而人类认识的目的和人类的需要，事实上具有整体联系。如果量子力学的数学形式体系需要物理解释，那么大数据也是如此。认为大数据时代我们理解世界不再需要建立在假设的基础上，无疑只是就人们的实务性信息活动而言，根本没有顾及人们更高层次的需要。

在创构过程中，越是高层次的创构，越是建立在“为什么”的基础之上，越是基于“为什么”的新的“是什么”。而且，这个“为什么”不仅关涉客观存在，而且与主观需要密切相关。没有对客观对象为什么这样存在，为什么以这种方式存在的因果性了解；没有对人为什么有这样的需要及这样的需要发展的理解，创构就既没有根据和基础，更没有方向和前景。我们甚至都不能理解创构物是什么，越是深度开发人类需要的创构物，越是如此。从技术上说，创构是通过组合操作，生产出满足表现出来的需要的过程。而从哲学上说，创构则建立在对客观对象的存在及其方式和人的需要及其发展的因果性认识基础之上。创构正是在这种“为什么”的基础上，生产出新的“是什么”。当创构成果成为新的“是什么”时，创构事实上是在既存东西的“是什么”到“为什么”，再到作为创构成果的新的“是什么”的过程。因素的表现组合只需根据大数据的量的分析，但所能满足的也只是人的表现需要，要真正使创构活动具有必不可少的基础，就必须具有相应的充分因果根据。不仅数据物化过程中的创构活动必须这样，而且物数据化基础上的数据分析也必须建立在相

应的因果基础之上。

### 三、大数据分析的因果性基础

把原因看作因素相互作用过程的描述，不仅可以揭示相关关系的因果性根基和数据物化的因果性根据，而且还将展示大数据分析的因果性基础。

#### （一）数据分析的因素分析性质

大数据条件使对结果的预期有了因素根据，而对结果的预期就是对因素以及它们的相互作用及其方式的把握。大数据时代之所以从难以捉摸的因果关系转而关注变量之间的相关关系，就因为在大数据条件下，在创构活动中，因果关系主要以相关关系的方式表现出来，而我们基于大数据的需要又能够应用因素分析的方法，通过相关关系得到满足。作为失去与因果关联的量化关系，相关关系不仅只能陷入真实相关和虚假相关的泥淖，得不到必要的说明，而且由此只能得到因果概念的传统理解。而作为因果派生关系，特别是作为创构时代因果关系的将来时形式，相关关系则不仅得到清晰表述，而且因素关系本身就表明大数据分析的因果基础就是因素分析。因为在因果关系的整体结构中，相对于其他因素相互作用过程，结果间关系也是因素关系。因此因素相关不仅是创构的前提，而且是数据分析的因果基础，即使在小数据时代也是如此。

无论是大数据时代还是小数据时代，因素关系都是一样存在，一样有用的，只是在描述性认识中，由于在相互作用过程中成了既定的因果关系，由于只是过去的关系，因而只能通过结果追溯原因，因素关系相对意义不像因果关系那么大；而在创构条件下，由于是从因素寻索结果，因素关系就显得至关重要。大数据时代，相关关系之所以“着实出色”，就因为所涉及的相关关系主要是因素关系，而因素分析不仅是一种具体的分析方法，在重新刻画的因果结构中，还具有作为数据分析因果基础的重要地位。

在大数据条件下，相关关系之所以如此重要，事实上正是因为发现相关性就可以预测因素相互作用的可能方式，就可以不同程度地预测未来结果。由此可见，说相关关系比因果关系更重要，应当是有条件的，那就是在大数据基础上达到相关性认识就足以满足需要。笼统地说“相关关系比因果关系更重要”，至少是以偏概全的结果。只是在对事物未来进程控制的意义上，相关关系比因果关系更有价值。而这种价值就来自与之相联系的特定潜在结果，这里的基础仍然是因果性。正是在这个意义上，因素分析集中体现了大数据相关分析的因果基础。

作为大数据相关分析因果基础的集中体现，因素分析的重要意义可以从定量分析的因素关系基础更清楚地看到。

## （二）定量分析的因果关系根基

由于从一般意义上说，任何事物现象都是相关的，因此从质的角度看，对于这种意义上的相关关系，定性研究就会遇到根本性的困难。定性研究不能对间接相关甚至直接相关关系进行操作性处理，因为越是间接相关，关系越是弥散的。但从量的角度看，如果对这种相关关系进行定量研究，就可以根据相关强弱程度和正负性质，进行定量操作。因而这种相关关系不仅很有意义，而且在大数据基础上，可以建立起一种具有操作性的处理方式。在大数据中，这种定量操作处理甚至使量具有特殊意义。

在信息时代，尤其是在大数据基础上，量具有与以往不同的意义。因为对我们的认识目的和实践需要来说，只有达到一定量的相关，达到一定相关程度，才具有认识和实践意义。因此，有没有和有多少就成了一个与传统哲学不同的问题。

大数据所涉及的都是量的关系，因而具有量的把握的特征：那就是一个数量和别的数量的相关关系，已经形成了一种具有量的确定性的联系。有了这种量的确定性联系，就能据此得到一些确定的结果，尽管可能只是信息而与物能没有直接联系。而在信息时代，更重要的结果恰恰正是信息。在大数据时代凸显的相关关系与因果关系的处理把握上，更根本的是信息因素和物能因素的区别。“大数据具有与世界 1 和世界 2 平起平坐的地位”，<sup>①</sup>也正是信息的这种基础性地位，使量的确定性具有越来越重要的意义。有了这种量的确定性联系，就有了可以通过相关因素的组合得到所想要结果的定量方式方法。有了量化基础和数学手段的运用，从而在定量把握的基础上，就可以通过相关因素的组合和安排，得到人们所需要的结果，甚至可以得到开发更深层次需要的创造性成果，包括满足人们需要的方式。

在大数据中，定量研究之所以有这么重要的意义，有两个层次的原因。首先，作为因果派生关系，大数据相关关系使全数据定量分析成为可能。相关关系提供了一个把握对象的量化基础，正是采用数学方法分析相关关系，才使因果性研究从因果关系推进到因素关系，从而可以进行定量研究。因素关系和因果关系的一个重要区别，在于传统的因果分析不能完全用定量分析方法，而因素分析则可以完全是定量分析。在因果关系中，数学方法的使用非常有限，但在相关关系处理中，数学方法的使用却不仅极为广泛，而且非常有效。相关关系之所以有时变得比因果关系更重要，也因为相关关系不像因果关系那么难以捉摸，而是可以用量的方式更好地分析事物之间的关系，特别是因素之间的相关关系，因为因素可以涉及人，涉及创构活动的需要及其发展的关系，从而使定量把握具有特殊价值。其次，正是作为因果派生关系，使通过全数据定量分析把握对象成为可能。只有作为因果派生关系，相

<sup>①</sup> 吕乃基：《大数据与认识论》，《中国软科学》2014 年第 9 期。

关关系才可能通过量化把握具有因果关系的对象。相对而言，因果关系更适用于质的把握，因而在定性研究上有优势；而相关关系则更适用于量的把握，因而在定量研究上有优势。通过因果派生关系即相关关系把握因果关系，虽然失去了必然性和方向性的优势，但是能够通过量的强度和正负值更有效地定量把握对象的因果关系，只是对于从相关关系到因果关系的推断要进行更多的深层次研究。由于在相关关系中，因素和结果之间的关系是对因果关系最直接的反映，大数据基础上的因果推断必须以相关关系的因果派生类型为根据。要进行正确的因果推断，必须区分不同类型的相关关系，用因素分析模型排除因素间和结果间的相关关系，得到因素和结果间的相关关系，这样才可能用统计的方法定量地把握对象的因果关系。

由此可见，相关关系意味着对事物过程的定量描述模型；而因果关系则意味着对事物过程的定性描述模型，二者都不可或缺。在复杂性系统和心灵哲学的身心关系研究中，可以感觉到来自两个相对方面的需要。以相关关系取代因果关系是短视；而以因果关系排斥相关关系则是固步。只有二者内在结合，才是描述事物过程的更合理方式，才能有对因果关系的全时态把握。原因的形成就是潜在因果关系的现实化，也就是从因素关系到因果关系。因果关系从潜在到现实，正是因素进入相互作用的过程。这个过程对人类活动之所以重要，不仅因为那是从一个由相关因素构成的可能性空间，通过选择满足自己需要的过程，更因为这是一个可以运用数学方法，通过定量分析进行创构的过程。因素进入相互作用前后的不同，说明了对作为一个过程的因果关系进行不同阶段研究和分析的重要性，这正是定量分析的因素关系基础，它是大数据分析的因果根据所在。

### （三）相关分析的因果派生依据

作为主要反映相关性的关系体系，大数据给人远离作为其根基的因果基础之感。表面上看，大数据所呈现的相关关系比因果关系更简单，但在相关关系之下，大数据却具有另一种复杂性。一方面，“就像宇宙大部分都是真空，大多数数据都只是噪声。”另一方面，“离开具体情境，数据一无所用。”事实证明，人们常常“将相关关系误为因果关系，将噪声误为信号”。<sup>①</sup>个中原因很多，既有大数据本身的，也有人们关于因果关系和相关关系的主观感知方面的。

埃丝特·戴森（Esther Dyson）曾谈到，人们对有关足球或赌博中统计性的理解水平让人惊讶，因此自然也就希望，“人们对相关关系和因果关系的区别同样有一种天生的理解。”<sup>②</sup>但实际上，不仅人们对相关关系和因果关系的直觉机制很不可

① Nate Silver, *The Signal and the Noise: Why Most Predictions Fail but Some Don't*, New York: The Penguin Press, 2012, pp.267,271,400.

② Esther Dyson, "Genetic Information," *Nature*, vol.455, no.7209, 2008, p.9.

靠，而且大数据相关关系所掩盖的实际内容也极为复杂。有别于“数据会自己说话”这种表层感觉，很多大数据研究者明确指出，“作为一堆海量原始信息，大数据不是自明的。”<sup>①</sup> 美国“数据和社会研究所”的创立者丹雅·博伊德（Danah Boyd）不仅认为大数据不是自明的，而且认为数据本身没有确定性。她的态度非常明确：“数字自己会发声吗？我认为，回答是一个响亮的‘否’”。在她看来，不仅数据自己不会说话，而且“海量数据可以提供辐射到所有方向的关联”。<sup>②</sup> 因此，一方面，大数据分析和操作与人们的主观选择有关；另一方面，人们对特定相关关系的理解不可避免地具有在前的理论前提。纽约大学教授丽莎·吉特尔曼（Lisa Gitelman）致力于文化数据研究，她甚至认为原始数据是一个矛盾体，对于数据的想象，每一个学科和学术机构都有自己的规范和标准。<sup>③</sup> 这正是人们有这样一种感叹的原因：“在数据中，将相关关系误当因果关系并找到误导性模型，那太容易了。”<sup>④</sup> 由此可见，大数据所呈现的相关关系远不是像表层操作中那么简单，以往关于外部世界认识的全部复杂性，都转移到了我们自己的理论根据，从而关系到大数据相关分析的因果性基础。在实际运用中，“大数据浮夸”正是由此导致的典型现象。这与大数据研究的真正问题所在密切相关，而所有这些都指向大数据的因果根基，指向大数据相关分析的理论基础。

对因素进入相互作用前的相关关系进行分析固然重要，但在因素分析的基础上再深化到因素相互作用过程的原因及因果分析，仍然在人类认识的深化中具有更为重要的地位。关于这一点，已为大数据研究中审慎而具有远见的专家所认识。内特·西尔弗（Nate Silver）就认为，“在大数据时代，这样的说法越来越司空见惯：有了这么多信息，谁还需要理论？但对预测来说，这样的态度绝对是错误的，尤其是在像经济那样的领域，数据那么杂乱。有了理论或至少关于其根本原因的某种更深入思考，统计推理就要更可靠得多。”<sup>⑤</sup> 而在大数据中越来越重要的因果推断，则不仅对因果性理解提出了更高要求，而且本身就表明了数据分析的因果旨归。在大数据中，因果推断即通过数据所反映的相关关系还原因果关系。这一因果关系还原过程是将数据变量之间的相关关系还原为因素相互作用过程，典型的如将数据关系还

① David Bollier, *The Promise and Peril of Big Data*, p.13.

② Danah Boyd and Kate Crawford, “Six Provocations for Big Data,” *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, Oxford: Oxford Internet Institute, 2011, pp.2,4.

③ 参见 Danah Boyd, *Six Provocations for Big Data*, p.5.

④ Andrew McAfee and Erik Brynjolfsson, “Big Data: The Management Revolution,” *Harvard Business Review*, October 2012, p.68.

⑤ Nate Silver, *The Signal and the Noise: Why Most Predictions Fail but Some Don't*, pp.211-212.

原为物的关系，也就是通过变量之间的关系认识因素相互作用关系，通过数据关系认识物的关系。作为因果检验模型，格兰杰因果检验是有效的，只是这一模型所检验出的不是原因和结果之间的关系，而是因素和结果之间的关系。所测出的关系强度也只是该因素在作为原因的因素相互作用过程中的作用强度。而通过变量之间的相关关系强度，只能检验出特定因素与结果之间的因（素和结）果关联，而不是真正意义上的因果关系。任何根据变量关系检验因果关系都只能检验出相关关系，而且只是检验出因素和结果间相关关系。当我们不满足于对大数据的相关因素分析，试图从“是什么”深入到“为什么”时，就必须进入到因果关系的深层次研究。

重新刻画的因果概念使我们看到，在大数据时代，因果关系不仅不会退居次于相关关系的地位，而且相反，作为因果派生关系，相关关系的当然基础是因果关系。作为因素相互作用过程的确定性关系，因果性不仅在更深层次关系到大数据的根基，而且关系到大数据更深层次的哲学内涵。

## 四、大数据的深层哲学内涵

作为与人类生存方式密切相关的新的存在方式，大数据提供了充分展开因果关系至其派生的相关关系的基础，具有深刻丰富的哲学内涵。基于其因果关系基础及作为因果派生关系的相关关系，在更深层次反思大数据的哲学内涵，无论对于大数据研究还是哲学发展，都当具有重要意义。

### （一）因素关系的未来空间凸显创构认识论

大数据提供了相对于潜在结果的因素关系空间，从而拓展了我们关于因果关系的理解，使人类的创构活动空前凸显，并以镜像方式将描述性认识和创构性认识呈现在人们面前。大数据所凸显的创构性认识，不仅意味着数据物化的认识论基础，而且意味着认识论的一个重要发展，具有深层哲学内涵。

关于创构活动的哲学反思，将形成一种不同的认识论。由于人工智能的发展，在关于认知系统设计的研究中，人们将认识论兴趣初步分成“程式认识论”和“描述认识论”。程式认识论旨在研究如何建立认知系统，而描述认识论则旨在研究如何描述系统运行时的功能。<sup>①</sup>在程式认识论基础上，虚拟技术使人们对这一认识论维度有更深入的理解，而大数据则给我们提供了进一步深入理解这种认识论的

① 参见 John Pollock, “Procedural Epistemology,” in Terrell Ward Bynum and James H. Moor, eds., *The Digital Phoenix: How Computers Are Changing Philosophy*, Oxford: Blackwell Publishers, Ltd., 1998, p.18.



基础。在认知系统设计的研究中,由于首先是创建一个认知系统,然后再描述其运行功能,这在认识论上与描述一个自然对象等价。在这个意义上,描述认识论是关于如何描述系统运行功能的,因此程式认识论在先,描述认识论在后。首先是如何建立认知系统的问题,然后才是怎样描述系统运行功能的问题。而在以往的认识论研究中,则主要是如何描述既存对象,因而是更广义的描述认识论。与“描述认识论”相对应,“创构认识论”更符合虚拟技术条件下,大数据出现之后的信息时代的性质。

因此,描述认识论是研究人类认识中以符号、语词和模型等摹写既存对象的性质、结构、前提和基础、发生和发展过程及其规律等的理论;而创构认识论则是研究人类认识中创设从未存在过的感性对象的性质、结构、前提和基础、发生和发展过程及其规律等的理论。描述认识论更具求真的特质,创构认识论则在因果根据的基础上,更具求善求美的特质。

描述认识论和创构认识论的明显分化,使得描述认识论意义上的因果关系,与创构认识论意义上的因果关系关联了起来,从而提供了在描述和创构之间深化因果关系理解的现实空间。

在描述认识论中,对象是因果关系,因为客观事实都是因素关系凝固的产物;而在创构认识论中,对象则是因素关系,因为可操作的对象必须是没有被凝固为因果关系的因素关系。正因为如此,创构认识论和描述认识论的不同,首先在于描述认识论主要基于因果关系,而创构认识论则主要基于作为因果派生关系的因素间相关关系。

创构认识论和描述认识论的另一重要不同,甚至涉及作为哲学核心内容之一的“意义”。在描述认识论中,因果关系的意义在于它与对象本身,越是反映对象根本性质的因果关系,意义越大;而在创构认识论中,因素相关的意义则取决于我们所要达到的目的,也就是取决于我们想象中的创构物。因此,与描述认识论中的情景不同,在创构认识论中自然就“不再强调意义的真实来源,而是强调意义如何被生产”。<sup>①</sup>前者是以事物存在和运动的“真实”为标准,而后者则是以满足我们实践的需要和达到认识的目的为标准的;前者的检验主要看是不是真实,后者的检验则主要看是不是能满足或开发人的需要,以及满足或开发人的需要到什么程度。

创构和描述、创构认识论和描述认识论一方面具有很大不同,另一方面又彼此内在相关。从描述认识论主要认识既存事物,到根据大数据提供的相关因素去建构新的事物,这意味着描述认识论意义上的因果性主要是由果溯因,而创构认识论意义上的因果性则更多是由因(素)构果,在一定意义上,这可以说是一个时代的转

<sup>①</sup> 卞友江:《“大数据”概念考辨》,《新闻研究导刊》2013年第5期。

换。由于相关关系的意义主要在未来而不是过去，其价值主要在于以一定的方式进入相互作用能得到一定的结果，因此对于未来结果的预测，甚至某种程度的把控，意义重大。一方面，未来是我们创构的；另一方面，只有在创构中，才可能真正有效地预测未来。这正是从由果溯因到由因（素）构果的转换，相关关系落实到因果关系之上的结果。

在大数据中，由于涉及创构，因素间相关关系的确拥有更好的前景。而作为因果派生关系，由于具有因果根基，由于具有量的把握的有效性，整个大数据相关关系则不仅使我们对因果关系的理解扩展到因素相关，而且量化的因素分析不仅作为方法，而且作为整体把握的手段，在哲学上具有重要后果。

## （二）因果派生关系的全数据定量分析呈现量的整体把握

把原因看作因素相互作用过程的描述，不仅可以揭示创构活动的认识论根据，因素关系本身就表明大数据相关分析就是具有方法论意义的因素分析，而且还将为大数据分析提供方法论基础。而通过量化世界，以定量的方式把握世界，则本身就具有重要哲学内涵。

当“量化一切”被视为“数据化的核心”，人们提出了“量化世界”的观念。在大数据基础上，数据化似乎使量化世界的理想离现实已经不远。在大数据中，物数据化使人们对世界事物的整体把握呈现全新的前景。由于物数据化的结果是作为因果派生关系的相关关系，因此为量的整体把握奠定了基础。

在哲学中，一个至关重要的基本问题，就是通过思辨达到对对象的抽象整体性把握。这种抽象普遍性虽然不意味着对世界的终极把握，但为具体事物的认识提供了必不可少的整体观照。而在大数据时代，作为因果派生关系的相关关系却给量的整体把握提供了一个重要基础，这是与小数据时代完全不同的，正所谓小样本时代“以小见大”，大数据时代则“以大见小”。大数据第一次使我们可以不只是通过抽象普遍性，而是在量的方面达到对一定对象的整体把握。这种量的整体把握及其重要性和必要性，早在量子力学形式体系中就得到了典型体现。只是由于人们不能从量子力学数学形式体系得到世界的物理图景，这一体现一直远离公众意识而没有发生应有的影响。与量子力学相比，大数据为我们提供了一个感性得多的场域。这是大数据与量子力学数学形式体系的相似之处，也是为什么大数据着实出色的另一个重要原因。

在量子理论中，我们一直在实证地应用其数学形式体系，但是我们也始终在寻求量子力学的理论解释，人类不会就停留在量子力学的数学形式体系把握上。大数据领域虽然与量子领域极为不同，但也可以由此看到，在大数据时代，只问“是什么”，不问“为什么”，也只是人们进入深层次思考之前的结果。因此大数据一定会问“为什么”，只是追问的目的和方式可能跟量子领域不一样。

在量子力学中,追问“为什么”的目的是在量的把握的基础上进一步达到对经验对象的质的整体把握;而在大数据领域,追问“为什么”的目的则是在此基础上创构出新的“是什么”。

大数据从量的方面建立起具体个别性的整体,为哲学的整体把握提供了新的手段,这种量的具体个别性的整体,跟抽象普遍性的整体把握具有不同的性质。由于通过思辨进行,抽象普遍性整体把握的不确定性是非常明显的。这种不确定不是与逻辑的严密性有关,而是源自逻辑与经验之间的契合性。而量的具体个别性的整体把握,却具有全数据的确定性。虽然一方面,大数据的大部分信息是噪音;但另一方面,我们又能够通过具体个别性全体的把握得到一个全数据意义上的整体性认识。大数据具有量化的整体性,是一个量化的整体,这使人们对世界事物的整体把握从质的抽象把握扩展到量的具体把握。通过大数据把握对象,具有与质的整体把握不同的性质。

就整体把握而言,质是对类的把握,但只是对具体事物的抽象把握;量则是对个别事物的具体把握,但只是量的方面的把握。在抽象普遍性观照之下,这种量的把握往往有比质的把握更重要的一面。

我们通过抽象普遍性达到的整体性,在哲学中具有重要地位,但同时也存在致思方向和把握方式方面的严重问题。抽象概括建立在有限归纳的基础之上,大数据则包括所有具体个别对象。即使描述都必须以某种规定为前提,在这些情况下通常都具有明显的主体对于信息的选择性,而大数据则在某种程度上趋向全信息。大数据使人们在分析某些现象时,可以处理海量的相关数据,而不再需要通过随机采样。大数据时代信息分析的这一转变,毫无疑问具有根本性。因为大数据基础上的信息分析不仅不涉及抽样,而且由于所处理的都是具体的数据,还可以不通过抽象就能对一定对象有一个整体把握,而且是量的具体把握。

在大数据中,与其说“量变已导致质变”,不如说,由于大数据量的特征,在那里,量和质的关系已悄然发生变化;与其说数据化“给予我们以量化的、分析的方式描绘世界图景的手段”,<sup>①</sup>不如说由于大数据的实践特质,使我们在世界图景的描绘中,认识和实践这两个基本面,已变得更为对称。而这种对称化的结果之一,就是从对抽象普遍性的过度倚重,走向与具体个别性的全面平衡。

在大数据领域,由于有时候量比质更为重要,因此量的全体把握对于实践具有特别重要的意义。在大数据时代,量比质之所以有时候具有更重要的意义,就因为因果关系量化为相关关系,而大数据本身就是量的关系全体。量化的全体虽然不能全面反映对象,但不仅能看出对我们具有实践意义的相关性,从而

<sup>①</sup> 参见 Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution that Will Transform How We Live, Work and Think*, pp.11,6,10,97.

预测未来的可能结果，做出有用的具体趋势性预测，而且能通过因果推断把握对象的因果关系。更为重要的是，我们总是采用更能满足需要的相关因素和因素相关，通过相互作用生成我们所需要的结果，达到我们的认识目的，满足我们的实践需要。现在开始流行的定制，就是通过大数据这个全体，观照个别顾客具体需求的有效方式。正是大数据的使用和人的需要的特殊关联，使大数据具有另一重要深层哲学内涵。

### （三）因果关系从描述到创构彰显哲学以满足人的需要为最终目的

大数据的另一个深层哲学内涵，则是人作为重要因素，可以通过调控其他因素，形成自己的预期结果，以满足自己的需要，因此，大数据以及在其基础之上的活动，使以人的需要为出发点，以满足这种需要为最终目的，在哲学中彰显出来，从而使当代哲学与传统形而上学的关系更为清晰。

传统形而上学从抽象普遍性出发，总是自觉不自觉地把抽象普遍性达到的最后的本质当作哲学的终极追寻。哲学在马克思那里，已经完全回到了感性实践。回到现实生活实践，归根到底就是回到人的需要，回到人的需要这一出发点，以人的需要为出发点，以满足人的需要为最终目的。

对既定存在对象的描述性认识，我们必须尽可能保持认识的客观性，而在创构活动中，面对海量信息，人们则主要从自己的需要出发，采用对人有益、有用，能满足人的需要的信息。在大数据时代，无论创构什么，都是以人的需要为出发点，而不是以追寻现象背后的终极原因甚至绝对的真为终极目的。因果关系是描述模型，而不是描述的对象。正因为如此，在大数据时代的信息分析中，发生了另一个重要转变：从热衷于精确度追求，转向由不精确数据反映的更小几率得到大趋势规律。这与认识目的和实践需要有关，并不意味着客观性的丧失，更并不意味着对因果关系理论本身意义的否定。

毫无疑问，“在很多情况下，相关关系自身极为有用”，但“运行数字和找到相关关系绝对不够”，因为“问题是，在分析这些数据集合的一开始，我们就必须对某些真正的核心理论作出决定性选择”。<sup>①</sup>此外，正是理论使数据及其使用和人的需要构成一个整体，赋予数据以意义。事实上，在理性场域，即使否定因果性也必须依赖因果观念，这意味着，即使否定理论本身也必须依靠另一种理论，而这样一来，必定陷入理性的自我矛盾之中。理论不仅必不可少，没有理论，数据毫无价值，甚至没有意义；而且，就像描述性认识中的描述和解释，创构性认识中的数据和理论互不可缺。一方面，“在实践中，理论和数据相互支撑。那不是关于数据相关关系对阵理论的问题。数据相关关系的使用

<sup>①</sup> David Bollier, *The Promise and Peril of Big Data*, pp.16,6,17.

使人们可以检验和精炼理论。”<sup>①</sup> 另一方面，即使观察都渗透了理论，数据的理解及其与人类需要的关联等，总是必须在一定的理论基础上进行，只是不要反过来，以理论本身为终极追寻。

大数据不仅意味着不以抽象普遍性为终极追寻，甚至也是不以终极因果性为最终追寻的最好例子。大数据使我们看到抽象普遍性追寻的意义和价值，同时又不会迷失于对它的终极追寻。就像抽象普遍性一样，相关性更不是终极目的本身，它的价值在于其所晓示的相关关系，只是必须理解相关关系的因果派生关系性质，从而使其落实到因果性根基之上。在因果性根基之上，作为因果派生关系的相关关系使因果关系本身研究的重要进程得以充分呈现。

因果关系的研究经历了一个从实在论因果观到描述论因果观的发展。实在论因果观陷于终极原因不能自拔，因此在休谟和康德之后，无论在哲学还是在科学中，无论在量子力学还是统计学和计量经济学等学科中，因果观念的实在论研究都发生了描述转向。而在大数据时代，从描述扩展到创构，因果关系的研究由主要对既存对象的描述，扩展到对潜在对象的创构，从而由于涉及创构而涉及意义的创生，涉及与人的需要的直接关联。正是大数据所推进的因果关系研究，最典型地晓示了以人的需要为出发点，以人的需要的满足为最终目的的哲学本性。

〔责任编辑：莫 斌 责任编审：柯锦华〕

---

<sup>①</sup> David Bollier, *The Promise and Peril of Big Data*, p.7.

## **ABSTRACTS**

**(1) The Meaning of the Scholastic Turn in Marx' s Later Years in Terms of Intellectual History** *Chen Zhonghe • 4 •*

The views that the scholastic turn in the later Marx was a turning from a universal to a particularistic view of history lacks any insight into the essence of the change in his ideas. In Marx' s youth, his ideas on society and history were mainly rooted in his study of Western European capitalism; although, to a certain extent, expressed as predictive world historical consciousness, they were essentially based on the historical consciousness of Western nations in the industrial age. Wittfogel firmly believed that scholastic turn in Marx' s later years represented a shift from a "universalist view of history" or "linear view of social development" to an accusation against the universalist concept of history, and reproached Marx for retreating "from truth." This understanding cannot withstand scientific criticism. In his later years, Marx acquired a broader view of world history through his study of ancient and Eastern societies, and thus made a true turn to and sublimation of world-historical consciousness. In his later years, Marx did not have time to fully unfold the significance of his world-historical ideas, but he did point out the direction for our further thinking at present about China' s future road and world-historical consciousness.

**(2) Causality in Big Data and Its Philosophical Connotations** *Wang Tiansi • 22 •*

In big data, datamation quantifies causal relations into relations between variables. This represents a gain in relationship strength and positive and negative properties, but a loss in terms of inevitability and direction. Big data correlation goes further in that it redefines the concept of causality: causality is a description of the process of interaction between factors, and what these correlations describe is causal derivative relationships. As causal derivative relationships, correlation is rooted in causality; as factor relations that have not yet entered the causal interaction process among factors and become fixed into causal relationships, correlation provides substantial space for factors to produce results, which constitutes the causal basis of data materialization; and as factor analysis, the causal derivative basis for the relevant quantitative analyses builds up the causal foundation for big data analytics. Big data' s clarification of causality reveals its underlying philosophical content. The future space of factor relationship highlights creative epistemology, the holistic grasp of quantities shown in total data

quantitative analysis of causal derivative relationships. From description to creation, causal relationships highlight the fact that the ultimate purpose of philosophy is to meet human needs.

### **(3) Human Capital and Housing Prices in Chinese Cities**

*Chen Binkai and Zhang Chuanchuan* • 43 •

The rise in housing prices in China has two central characteristics: it is non-stationary and heterogeneous. Using household micro-data, we analyzed the dynamic evolution and regional differences in housing prices in Chinese cities from the perspective of the growth of human capital. Our analysis shows that the expanded enrolment in higher education since 1999 and the increased scale and spatial concentration of human capital arising from the unequal distribution of higher education resources between regions are important factors in these two characteristics. For each one percent increase in the proportion of the population with higher education, urban housing prices will rise between 4.6% and 7.9%. This can explain some 12% to 20% or so of the increase in housing prices from 2002 to 2009. This comes about via the “direct effect,” the “spillover effect” and the “anticipation effect.” As the pace of the increase in human capital tends to slow down, the Chinese housing market has entered a “new normal” state. The policy of short-term stimulus to housing prices does nothing for the steady and sound development of the housing market.

### **(4) Partners, Systems and International Currencies: An International Political Foundation for the Rise of RMB**

*Li Wei* • 79 •

The use of international currencies is not simply a natural market phenomenon, but also depends on a sound international and political foundation. As it is shown by the historical lesson of four major international currencies, namely the British pound, the US dollar, the Euro and Japanese yen, a powerful currency partner network and an advantageous international institutional system are an important international political foundation for an international currency to build up a stable position. As a late-developing currency, the RMB will have to gain more monetary cooperative partners and create international institutions supporting the international use of the RMB, as well as having the right economic conditions, if it is to lay a firm international political foundation for RMB's rise as a major currency on the global stage. Whether we are establishing a partner network of or creating international currency institutions system for it is a challenge for the political leadership of an issuing country of international currency. Such political leadership has three dimensions: credibility based on reputation building, centripetal force based on profit supply and coercive force based on implementation of sanctions. Compared with others, the RMB as an international currency is still at the initial stages of its rise; it is thus even more necessary, in constructing its international political

• 205 •