

第五篇

元求索

第二十三章

基本求索的相互关系

在本章中，我首先提出在基础求索中存在一种原语**求索** (primitive interrogation)，然后进一步考察四种基础求索：**测量**、**测试**、**推理**和**评价**之间的关系。

23.1 原语求索：与**参考**进行比较

在第**五**，第**六**，第**七**和第**八**章，我分别定义了测量、测试、推理和评价。

在分析四种基础求索的基础上，可以找到一种原始的求索，即 将一个**对象**与一个参考对象进行比较。对于测量，**主体**会将任意对象的**量**与参考对象的**测量单位**进行比较。对于测试与**检验**，主体会运行**测试用例**，并将结果与检验标准的结果进行比较。检验标准定义了一个参考，用于比较。

推理依赖于**推理规则**。这些推理规则可以认为是推理的参考。同时，推理的结果只能通过测试来验证，而测试需要与参考（**检验基准**）进行比较。

不同对象的评价结果只有在相同的**评价基准**下才具有可比性。这里的评价基准正是一个参考。

23.2 四种求索的独特位置

测量、测试、评价和推理代表了四种基本的求索方式，通过它们，智能生命探索未知世界，并构建知识体系。

测量和测试回答“多少”，为对象的量赋值；评价回答“为什么”，揭示原因对象的**影响**或现象的原因；推理则是通过主体的**心智**活动产生新的命题和模型；通过与检验基准比较，测试负责检验对象或者对象关系的命题或者模型的有效性；综上所述，这四种求索共同构成了一个完整的认知循环——从观察和实验，建模，到验证¹。

¹范帆达为本段提供了最初的文本。

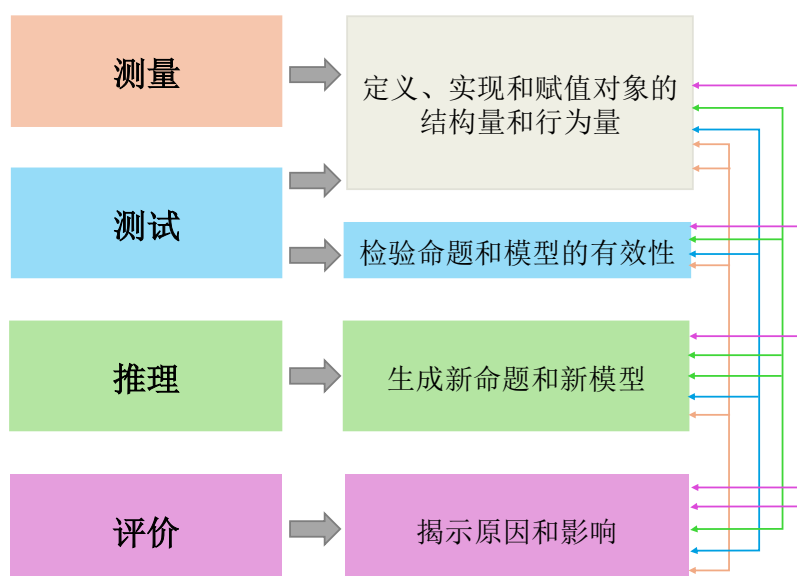


图 23.1: 四种求索的相互关系。

23.3 四种求索的相互关系

这四种求索的相互关系如图 23.1 所示²。

测量和测试为对象定义、实现和赋值结构量和行为量，测量、测试、推理和评价这四个基本求索依赖于测量和测试定义、实现和赋值结构量和行为量。

四个基础求索都依赖推理产生对象或对象关系的新命题或模型，但这些只能通过测试来检验。检验基准本身包含检验通过的命题和模型。在已有检验基准的基础上，基于推理也能产生新的检验基准。

评价依赖于测量、测试和推理来推测对象的影响或现象的原因，在此基础上可以生成命题或者模型，它们可以作为测试或者推理的输入。

23.4 总结

本章总结了原语求索：比较与四种基础求索在人类认知活动中的独特位置，同时讨论了四种求索的相互关系。

²高婉铃在和我及王磊讨论的基础上，绘制了该图。

第二十四章

评价学在求索里的应用

本章尝试将评价学应用在求索活动中。贡献者为康国新。

24.1 引言

在第 5.10 节, 我们定义了求索的基本量。本质上是将计量学应用于求索的各个阶段: 为测量、测试、推理和评价等基本的求索活动定义、实现和赋值量。

然而, 仅仅定义求索结果的量仍然不够。任何求索结果都不是孤立产生的, 而是在特定求索对象、求索条件、求索主体、求索方法和求索工具共同作用下形成的结果。因此, 当我们得到一个测量或测试值、检验结论、推理结论或评价结果时, 还必须进一步追问: 这个结果为什么如此? 哪些组件对结果产生了影响? 这种影响是直接的还是间接的? 如果可更换求索工具、改变求索条件、替换求索方法, 求索结果会如何变化? 这些问题的回答可以借助于评价学的方法。

本章阐述评价学基础方法在求索活动中的应用。求索包含测量、测试、推理与评价四类活动, 它们都会产生相应的求索结果。可以从两个角度探讨评价学在求索的应用, 一是将求索过程作为一个整体, 将某一类求索活动产生的求索结果作为原因对象, 使用评价学揭示该原因对象的影响。其次是把求索活动当做一个白盒, 将其内部组成结构作为原因对象, 揭示其对求索结果的影响。本章侧重第一个角度。

本章首先建立将求索结果作为评价对象的通用框架, 明确定义影响对象、混杂对象与影响机制; 随后以测量为例, 讨论评价学在求索活动中的应用; 最后进行本章小结。

24.2 通用框架

24.2.1 评价对象、影响对象与混杂对象

评价对象 (EO)。在本章中, 评价对象假设为某一类求索活动所产生的求索结果, 记为 $\hat{\theta}$ 。依据求索类型的不同, $\hat{\theta}$ 可以是测量或者测试得到的量值、检验结果、推理生成的新命题或估计量, 以及评价活动推断出的影响。根据前述定义, 求索结果可以是对象或者衍生对象, 完全可以成为评价对象; 当我们揭示求索结果的影响时, 求索结果即为原因对象。

从另外一个角度理解,测量、测试、推理和评价系统都是由求索对象、求索条件以及求索工具等组件组成的求索系统。求索系统的组成,例如求索工具,会对求索结果产生影响,也会通过求索结果间接影响其他对象。此时,求索工具就是原因对象,求索结果变成了影响对象。可见对象的角色是相对的,由评价的目标决定,而非对象本身所固有。以测量为例,我们既可评价测得的量值对模型的影响,也可评价测量工具对测得的量值的影响。前者把求索结果视为评价对象,后者把求索结果视为影响对象。

求索结果的一个关键隐藏量是它的真值 θ 。相对于真值,求索结果的偏差或者误差并不能被直接测量或者测试,只能通过推理得到。正是这一原因,评价求索结果的影响成为一个极具现实意义的课题:我们关心的不仅是数值本身,更是带有偏差或者误差的结果在被采纳时会引发的连锁反应,也就是影响。

影响对象 (AO)。求索结果一经产生,便会被主体用于构建命题、完善模型或指导决策。因此,求索结果的**直接影响对象**是直接消费该结果的命题、模型或决策。例如,测量结果的直接影响对象可能是将该测量结果作为自变量输入的参数估计模型或工程控制决策系统。

求索结果的影响会沿着影响路径进一步传导至**间接影响对象**,并最终触及利益相关方——即对该求索结果负有责任或持有切身利益的智能生命及其组成的系统。例如,某次测量结果的误差,通过模型传递至一项工程决策,最终影响到承担该决策后果的工程师、企业实体或社会公众。依照既有定义,在求索场景中,不同求索结果对利益相关方的影响差异对利益相关方的影响,即体现为该求索结果的价值。

混杂对象 (COs)。在现实条件下,下游的命题、模型或决策 (AO) 并不只受被评价的单一求索结果影响,还会受到其他混杂对象的共同作用。在评价求索结果的上下文中,典型的 COs 包括:进入同一下游模型的决策规则、主体所持有的领域先验知识、以及难以完全剔除的环境扰动。求索结果与混杂对象集合共同作用于影响对象,产生总体影响;评价的核心任务,正是从总体影响中剥离并量化该求索结果的真实影响。

24.2.2 影响机制:信息传递与误差传播

求索结果作为原因对象,通过影响机制对影响对象产生影响,一种典型的影响机制是信息传递与误差传播,可以通过观察或者**实验**等基础求索方法揭示。一个求索结果以其数值形态进入下游的命题、模型或决策;与此同时,它的误差或者偏差也被引入系统。下游对象通过特定的映射关系(函数、规则或物理机制)依赖于该结果,因此结果的偏离会沿着这一映射关系传播,最终转化为下游对象上可观察、可测量或可测试的差异。

形式上,假设影响对象的某个被测量为 $Z = g(\hat{\theta}, s_1, \dots, s_k)$, 其中 $\hat{\theta}$ 是被评价的求索结果 (EO), $s_1, \dots, s_k$ 是混杂对象集合 (COs) 对应的其他系统输入。则求索结果作为原因对象,对影响对象的影响可由灵敏度进行刻画:

$$\delta Z = \frac{\partial g}{\partial \hat{\theta}} \delta \hat{\theta} + \sum_{i=1}^k \frac{\partial g}{\partial s_i} \delta s_i. \quad (24.1)$$

公式中第一项代表求索结果对下游对象的直接影响,后续各项之和代表原因对象经由混杂对象对下游对象产生的间接影响。这一影响机制在数学形式上与不确定度传播一致,都是一阶灵敏度与各输入偏离之积的求和,但两者的视角恰好相反。不确定度传播把求索结果写成其自身输入的函数 $\hat{\theta} = f(x_1, \dots, x_n)$, 合成的是 $\hat{\theta}$ 的不确定度,此时求索结果位于映射的输出端,是影响对象。本章把影响对象的被测量作为影响对象,求索结果

为输入的函数 $Z = g(\hat{\theta}, \dots)$ ，传播的是 $\hat{\theta}$ 的偏离，此时求索结果位于映射的输入端，是原因对象。同样的一阶展开，求索结果的位置由输出端移到输入端，所刻画的对象也由求索结果的生成转为求索结果的影响。

需要强调的是，这一影响机制并不局限于线性可微的情形。当下游映射关系表现为测试中的判定规则、推理中的逻辑蕴涵或评价中的价值函数时，影响机制相应地表现为判定结果的翻转、命题真假的逆转或价值排序的重组。

24.2.3 现实世界系统、自包含系统与对象因果模型

评价任何求索结果，都必须严格遵循评价学基础流程：建立现实世界系统、隔离自包含系统、建模对象因果模型。

现实世界系统 (RWS)。在现实场景中，求索结果是被深度嵌入复杂场景中的，它与海量的其他结果、业务规则、辅助工具和环境因素相互交织，共同影响着主体的认知演进与决策行为。直接在现实世界系统中揭示某单一求索结果的真实影响极为困难，因为研究者难以准确划分原因对象、混杂对象和影响对象的边界，难以有效剔除无关对象，也无法随意操控复杂的求索条件。

自包含系统 (SCS)。为突破现实世界系统的困境，评价的关键在于隔离出自包含系统：

$$SCS = \{\hat{\theta}(\text{EO}), \text{AO}, \text{COs}\}. \quad (24.2)$$

该系统必须包含被评价的求索结果、其影响得以体现的最小显式对象，以及与之共同发生作用的关键混杂对象。构建自包含系统需严格遵循评价系统构建的四个原则：同一性、简单性、递进性和可观察性。在求索的评价上下文里，自包含系统通常通过受控的可重复求索条件来实现：即固定混杂对象集合（例如固化下游预测模型或决策规则），仅允许被评价的求索结果在其取值或偏离区间内变化，从而将求索结果的影响从总体影响中有效隔离。

对象因果模型 (OCM)。在构建的自包含系统之上，进一步建立对象因果模型，刻画评价对象、混杂对象和影响对象之间的关系，明确评价对象的影响如何通过影响机制作用于直接和间接影响对象。

24.3 测量的评价

问题背景

测量是指在具体求索条件下，将测量对象的量与测量标准规定的参考对象单位量进行比较，从而为测量对象的量赋值的过程。测量活动得到量值，之后该量值会被直接进入下游的模型、统计算法或工程控制决策中。因此，评价某一测量结果，旨在揭示这个客观上带有误差的量值在实际应用时所产生的真实影响。

现实世界系统

在现实世界中，同一个被测量往往需要进行反复校验，测得的量值随之进入形式各异的计算、报告与决策链条。这些链条彼此交织，叠加着仪器衰减、环境温度、操作人员习惯乃至下游算法截断误差的诸多复合影响。直接在现实世界系统中评价某次特定测量造

成了何种后果是极其困难的：我们既无法剥离其他混杂对象的影响，也难以低成本地复现测量条件。

自包含系统

为评价单次测量结果的影响，可行的方式是隔离出下述的自包含系统：评价对象（EO）是该测量结果 $\hat{y}$ ；直接影响对象（AO）是接收 $\hat{y}$ 作为输入的最小下游模型或决策模块——例如一个以 $\hat{y}$ 为自变量的参数估计算法、动力学模型或工业控制逻辑；混杂对象集合（COs）是同时进入该下游模型的其他辅助测量结果 $x_1, \dots, x_k$ 、下游模型的计算规则体系，以及作为基准参考的测量标准。

这一自包含系统通过受控的机制予以实现：固定下游模型与其他输入（COs），强制令测量结果在合理误差区间内进行微小扰动，并严密观测下游模型输出的连锁变化。这正贯彻了自包含系统的设计哲学——通过仅改变评价对象而控制其余混杂对象，从而将该评价对象的因果影响从总体影响中剥离出来。

对象因果模型

测量结果的对象因果模型建立在被测量与下游复合量之间的函数映射关系之上。设定下游量为 $Z = g(\hat{y}, x_1, \dots, x_k)$ 。则测量结果对下游量的直接影响计算为：

$$E_{\hat{y} \Rightarrow Z} = \frac{\partial g}{\partial \hat{y}} (\hat{y} - y), \quad (24.3)$$

其中 y 代表被测量的绝对真值， $\partial g / \partial \hat{y}$ 为该点的灵敏度系数。当下游量进一步干预某项宏观工程决策或触及利益相关方时，测量结果对其产生的衍生影响即转化为间接影响 $E_{\hat{y} \Rightarrow Z \rightsquigarrow \text{利益相关方}}$ 。

24.4 本章小结

本章将评价学方法应用于求索活动自身，构建了以求索结果为评价对象的统一评价框架。无论是测量、测试、推理还是评价的结果，均可作为评价对象。它们直接影响下游的系统、模型或决策；而环境、规则与底层工具等则构成混杂对象集合。求索结果偏离真值的初始误差，会沿逻辑链路传播，最终在影响对象上表现为可观察、可测量或可测试的差异。

在此框架下，评价任何求索结果均需遵循标准流程：界定现实世界系统、构建自包含系统、建立对象因果模型，进而精准剥离出该结果的真实影响及其对利益相关方的实际价值。