预惑

预测的时间性自反与本体论困境

金华 著 · 依王德生《SDE本体论》 · 发表于2026年7月28日

摘要

预测通常被理解为一种“向前看”的理性行为。然而，无论其技术手段如何精密，预测行为在本体论上是否必然陷入一种结构性的自我困惑？我们的回答是肯定的，并将其命名为“预惑”（proleptic perplexity）。所谓预惑，不是心理学层面上的“担心预测不准”，也不是指向未来之不确定性的认识论慨叹；它是预测行为在时间性自反结构中所必然卷入的这样一种困境：预测试图用“现在”的认知工具去抵达“未来”，却使得那个被抵达者不再是未来本身；并且，预测者永远无法区分，在被观察到的那个未来的生成过程中，有多少是由预测行为自身参与制造的。本文通过与默顿的自我实现预言、麦肯齐的操演性理论、哈金的循环效应、以及奥斯汀的言语行为理论的硬碰硬对话，论证预惑切开了上述理论装不下的新辨别面：它揭示的并非“预测如何改变现实”，而是“预测行为在定义自身时所必然卷入的时间方向上的自我取消”——即对未来的事先把握，必然取消那个被把握者的未来性。通过对2008年金融危机风险模型和预测性警务算法争议的深度个案分析，本文进一步追踪，预惑作为一种生成性矛盾，如何在制度运作中被系统性地驱离——不是被某个有决心的主体所逃避，而是被一整套技术-官僚装置所遮蔽、分散和转译。当代制度对预惑的驱离，其核心机制并非有意的欺骗，而是以精密化遮蔽本体论困境，以结构性回避取代行动者的存在性承担。当这种驱离在危机中瓦解时，被压抑的预惑将以公共合法性崩溃的形式猛烈爆发。在论文的结论部分，我们将阐明，预惑对于社会决策的意义不在于它可以被解决，而在于它要求预测的部署者与被预测者之间必须有一种被制度保护的、对预测本体论地位的公开争辩空间——不是为了提高预测的准确性，而是为了使“这个预测正在制造什么”这一问题永远不被卸下公共议程。

关键词

预惑；时间性自反；操演性；本体论困境；生成性矛盾

引言　一个被遮蔽的困惑

我们生活在一个被预测填满的时代。每日财经新闻的头条是央行对通胀的预测，手机应用推送明天下雨的概率，教育系统从小学阶段就开始用标准化测试预测一个孩子的“未来学业成就”，算法在你尚未意识到之前就已预测并推送你的下一件欲购之物。与此同时，这恰恰也是一个对预测充满焦虑与嘲讽的时代。2008年全球金融危机的集体预测失败、2016年美国总统大选几乎全部民调的失准、新冠大流行期间流行病学模型前后矛盾的修正——每当重大预测失灵发生，都会引发一轮对专家知识权威的公开羞辱。

标准的应对是技术性的：“我们的模型还需要更多数据，更复杂的算法。”其潜台词是：预测的问题在于它还不够好，而当它足够好时，一切都会解决。这种技术性回应的背后，潜伏着一个从未被严肃追问的预设：未来本身可以被预测。　　这里的“可以”不是技术可行性的判断，而是一种更根本的、关于未来之本体论地位的前反思默认：未来被当作一个“已经在那儿”、只是尚待我们前去认识的实体。问题仅仅在于我们的认识工具暂时还不够锋利。

然而，这一默认并非不言自明。早在二十世纪上半叶，弗兰克·奈特在其《风险、不确定性与利润》中就已经做出了一个决定性的区分：风险是可量化的未来可能状态，而不确定性是不可量化、概率分布未知的未来可能状态。奈特的洞见在于，真正的不确定性不是“我们还不知道概率”，而是“未来本身就不可度量”。但即使奈特，也仍然停留在这样一个假设之内：未来的确存在某些状态，无论我们是否能量化它们，它们总归“在那里”。他的批判所撼动的，是我们对未来的认知能力，而非未来作为认知对象的存在方式本身。

本文要追问的，正是这个更深一层的、关于“预测-未来”关系之本体论地位的信念。我们将论证：预测行为，在一切技术考量之前，在本体论上就必然陷入一种结构性的自我困惑。这种困惑不是预测操作层面上的“可能失败”，而是预测在定义自身的那个动作内部无法规避的自我反噬。它既不是奈特意义上的不确定性——因为不确定性指向的是未来状态的不可知，而这种困惑指向的是预测行为自身的不可定位：它不知道，自己所声称要描述的那个“未来”，在多大程度上是被它自己的宣告制造出来的、在多大程度上是原本就会发生的、而又在多大程度上永远无从知晓——因为那个可被用来比对的、预测未发生时的未来本貌，在任何存在时间性自反的系统中，根本就不存在。我们将这种困境命名为“预惑”（proleptic perplexity）。

这一困境并非书斋中的抽象思辨，它深嵌在当代社会最具体的技术与制度实践之中。2008年金融危机中，全球最精密的金融风险模型集体失效，不是因为它们不够复杂，而是因为它们所嵌入的制度体系正是被设计来系统性驱离预惑的——而驱离的代价，则是预惑在危机时刻以整个认知体制合法性蒸发的方式猛烈反噬。在预测性警务算法的争议中，一个被COMPAS系统标记为“高再犯风险”的人，他的未来在预测宣告的那一刻就已经被改变——更多的巡逻、更严密的监控、更早的逮捕，所有这些都使他更可能“兑现”那个预测。而算法的设计者与使用者却说：看，我们预测得很准。预惑在这里被封装进黑箱，被转译为一个纯技术性的“分类误差问题”——它不再表现为一种可以被感知、被追问的存在性困惑，而只是一组有待优化的准确率数字。

本文的论述将沿以下路径展开。在第一部分，我们将对预惑进行严格的理论锻造：不是通过给出一个先验的定义，而是通过让其与既有的、已经触及“预测-对象”反馈循环的理论传统进行硬碰硬的对话，逼出预惑不可还原的独特辨别面。这一部分的核心对话对象将包括默顿的自我实现预言、麦肯齐的操演性、哈金的循环效应，以及奥斯汀的言语行为理论。在第二部分，我们将以2008年金融危机中风险价值模型（VaR）的生成、运作与崩溃为深度个案，追踪预惑作为一种生成性矛盾，如何在特定历史条件下逼迫制度走向一种以“精密化遮蔽本体论困境”为核心的逃避形态——以及这种逃避为何不能简单归结为“系统的自我保护”，而有其独立的发生动力。在第三部分，我们将目光转向当代的算法预测，以美国司法系统中广泛使用的COMPAS风险评估算法为核心案例，剖析预惑在当代技术环境下如何被重塑——从一种行动者与未来之间的时间性困境，被转译为一个由数据-模型-输出闭环所封装的技术性“分类误差问题”。在结论部分，我们将阐明预惑这一分析视角的实践意涵：它要求我们在制度层面为“这个预测正在制造什么”保留一种无法被算法化、无法被外包的公开争辩空间。

一、“预惑”的生成与辨别

1.1 预惑的诞生地：预测行为的时间性自反

预惑的诞生地，是预测行为不可分离的一种特性：时间性自反性。这里所说的“自反”，不是社会学意义上的“反身性”——不是认知者回过头来审视自己的认知条件——而是时间方向上的因果回卷：一个关于未来的断言，在它被做出的那一刻，就已经成为那个未来得以生成的一股因果力量；而这股力量所参与制造的那个未来，又被回过来当作验证或否定该断言的依据。

让我们从一个具体场景入手，以锚定这一概念的直观所指。设想一位宏观经济学家，在某年第一季度做出预测：“年底通胀率将达到4%。”这个断言在逻辑上预设了两件事：第一，存在一个客观的未来事态（年底通胀率），它在那个时间点“在那里”等待被实现或观察；第二，现在的断言与未来的那个事态之间存在一种真值关系——断言或是真的（若通胀率确为4%），或是假的（若不是）。

现在，时间性自反性在第二点上发动了攻击。这位经济学家的预测一旦公开发布，便立即成为经济系统中的一个因果变量。消费者读到这则预测，提前囤积商品；企业据此调整定价策略；央行在看到通胀预期升温后，可能提前加息。所有这些行为，都改变了年底通胀率的实际走向。那么问题来了：当年底通胀率果真接近4%时，我们能说“预测正确”吗？这4%在多大程度上恰恰是预测行为本身制造出来的——若无此预测，通胀率本应只有2%？或者相反，央行的加息操作抑制了通胀，年底数据停在3%。我们能说“预测错误”吗？也许在预测未发生的情况下，通胀本会飙升至6%，央行的干预恰恰使结果趋近于预测值。更根本的是：那个可以用来比对的“预测未发生时的未来本貌”，在任何存在时间性自反的系统中，根本就不存在——因为它从未发生。

这种困境，不同于我们熟悉的风险与不确定性。风险是可量化的未来可能状态（如掷骰子的概率分布）；不确定性是不可量化的未来可能状态（如金融危机的概率分布未知）。但无论风险还是不确定性，都共享一个假设：未来存在某些状态，无论我们是否能认知它们，它们总归是“那里”的某些东西。预惑则不同。它的困境不在于“我们不知道未来是哪个状态”，而在于“我们无法知道：那个被观察到的未来状态，与我们的预测行为之间，因果账该怎么算”。这不是一种认知的空白，而是一种因果的不可分解。

让我们以医疗领域的一个更微观的场景，来让这一困境更可见。一位精神科医生对一位初诊患者做出预测：“根据症状模式，你很可能患有双相情感障碍，需要长期服用心境稳定剂。”这个预测当然不是凭空而来——它基于诊断标准、流行病学数据和医生的临床经验。但这个预测一旦被告知患者，它就已经开始改变患者的自我认知和行为轨迹。“我是双相患者”这一标签，可能使患者将原本正常范围内的情绪波动解释为“症状”，从而增加就医频次、强化病人角色、甚至在无意中调整自己的行为以使其符合诊断叙事。几年后，医生回看病例：“我的诊断是正确的——他的病程确实符合双相障碍的典型发展。”但医生永远无法知道：如果当初他给出的不是一个确定性的诊断标签，而是“你的情绪状态确实有些波动，我们一起来理解它”，这个人的生命历程是否会走向另一条路。

预惑的实质，正是这种“账算不清”的结构性处境：预测的行为已经不可逆转地卷入了它所描述的对象之中，而那个可以被用来做“干净比较”的控制组世界——一个预测从未发生、未来的面貌纯

粹地由非预测性力量决定的世界——在任何存在自反性的系统中都无法获得。它不是预测不准时的遗憾，它就是预测在构成自身时所同时制造的那个不可消解的盲区。

1.2 敌意最近邻：预惑与既有理论的硬碰撞

“预测可能改变被预测对象”——这一直觉并非本文首创。在社会科学与哲学中，至少四支理论谱系已从不同角度触及了此问题，其中一些与本文的核心案例直接重叠。要论证“预惑”的不可还原性，就必须与这些真正的敌意最近邻进行硬碰硬的对话——不是指出它们“没说到”，而是展示在什么条件下它们的分析停住了，而预惑在什么条件下继续推进，切出了一个它们装不下的新辨别面。

第一支：默顿的自我实现预言。 罗伯特·默顿在其1948年的经典论文中提出了一个至今仍有强大解释力的概念：当一个关于未来的虚假定义被广泛相信时，它会引发行为改变，使原本虚假的定义变为真实。一家运行良好的银行，若被传言“将破产”，储户蜂拥挤兑，银行果真破产——这是自我实现预言的范式案例。默顿的框架极富洞见，它精确地揭示了一种“预言参与制造自身之真”的社会因果机制。

但预惑与自我实现预言之间的差别，恰恰是默顿的框架所装不下的。自我实现预言需要一个干净的因果叙事起点：先有一个“假定义”，然后这个假定义通过改变行为而被做实。这个叙事之所以能成立，是因为在默顿的范式中，我们能够追溯那个“假定义”的来源（一个谣言、一个错误判断），并将它与事后被“做实”的结果进行比照。然而，在预惑的困境中，这样的干净叙事恰恰不可能。要判断一个预测是“正确的预测”还是“自我实现的预测”抑或“自我挫败的预测”，需要一个独立于该预测的控制组世界作为参照——但这个参照系在时间性自反的系统中不存在。预惑不是默顿机制的一个特例，而是默顿机制在本体论上的困境版本：它追问的不是“这个预测是否通过社会因果机制改变了现实”，而是“预测者何以可能知道自己正处在正确/自证/自败的哪一支上”——而答案是：他不可能知道。这不是暂时的无知，是结构性的不可知。默顿给我们的是一个社会因果机制，预惑追问的是这个机制之下的本体论处境。

第二支：麦肯齐的操演性。 在社会研究领域，唐纳德·麦肯齐在其《引擎而非照相机》一书中对金融模型的经典分析，构成了与本文VaR案例的最直接重叠。麦肯齐的核心判断是：金融模型并非像照相机那样“拍摄”一个预先存在的市场现实，而是像引擎那样参与制造它。以Black-Scholes期权定价公式为例：该公式在1970年代被提出后，交易员开始用它来为期权定价；随着使用者的增加，市场价格逐渐收敛于公式所给出的理论值；公式因此变得越来越“准确”——但这不是因为它发现了市场的本质规律，而是因为它重塑了市场的运作方式。操演性理论在金融社会学中已被广泛接受，它有力地瓦解了“模型只是描述现实”的发现学预设。

预惑与操演性共享一个起点：对“模型作为中性描述”的批判。但二者的焦点从这一共同起点处分岔。麦肯齐的操演性关心的是“模型做了什么”：它如何改变了交易实践、市场价格、制度安排；它的分析单位是市场层面的“操演-反操演”互动。预惑关心的则是一个操演性装不下的问题：预测行为的时间方向上的自我取消。 操演性问：“模型如何影响了世界？”预惑问：“当模型影响了世界之后，‘模型是准确的’这句话，还能在什么意义上成立？”操演性可以详尽地描述Black-Scholes如何从“不准确”变为“准确”，但它不追问这个“准确”本身的含义是否已经被掏空——因为那个可资比对的、未经模型干预的市场本貌，已经永远消失了。更精确地区分是：操演性分析对于“模型塑造了市场”这一判断与对于“这个模型是否准确描述了市场”这一判断之间的关系，持的是两可态度；

预惑则切入了这一关系本身——它指出，在时间性自反的系统中，“塑造”与“描述”这两个动作的界线在本体论上就是模糊的，无法被分解。每一个“描述”都已经是“塑造”的一部分，而“塑造”的后果又被回过来当作“描述”之准确性的证据。操演性理论没有精确捕捉这种循环中“未来被取消”的特有困境：一个预测用现在的概念工具去抵达未来，但抵达的同时就使那个未来不再是原本可能发生的未来——它的“未来性”（作为尚未到来者）被预测行为本身取消了。

第三支：哈金的循环效应。　伊恩·哈金在其对“构造人”与“人类种类的循环效应”的持续研究中，发展出了与预惑的另一个侧面直接相关的论点——尤其与本文第三章COMPAS案例高度亲和。哈金的论证是：当专家系统创造出某种分类（如“多重人格障碍”“少年犯”“高再犯风险者”）并将之应用于真实的人时，被分类者会意识到自己正被如此分类；这种意识会改变他们的自我理解与行为方式；被改变的行为反过来为分类的“真实性”提供了新的证据——从而形成一个“分类↔行为”的循环。哈金称此为人与自然种类之间的根本差异：夸克不在乎物理学家怎么叫它，但人会在乎社会怎么分类他。

预惑与哈金的循环效应有不可否认的亲缘。二者共享一个起点：对“分类如何参与制造被分类者”的敏感。但预惑在此之上追加了一个哈金未独立处理的维度：不可逆的时间性与因果账的无法结算。　哈金的分析焦点是分类概念与社会现实之间的互动循环，但他并不特地区分“预测未来”与“描述现在”这两种言语行为的本体论差异。预惑则将这一差异推到前台：预测不是对既有状态的分类（如“你是多重人格”），而是对尚未发生之事的断言（“你将会再犯”）。当这个“预分类”改变了一个人的未来轨迹时——他被更密切地监控、更早地被逮捕、更不可能获得假释——事后用以验证或否定这个预测的“再犯数据”，本身就沾染了预测行为的因果痕迹。我们可以说出这个人的“再犯被捕率”是多少，但我们永远不知道在另一个没有使用这套算法的世界里，他的“实际再犯率”会是多少——因为那个世界从未发生，也永远无法发生。哈金的循环效应解释了“分类”如何制造现实，但它并不内在地追问这样一个问题：在“分类”与“现实”的循环已经闭合之后，那个声称“我只是在做风险评估”的预测者，被他自己的预测置于一种什么样的存在性位置？　预惑追问的恰是这个问题——而且是就“预测”这一特定的、时间上指向未来的分类行为来追问的。

第四支：奥斯汀的言语行为理论。　约翰·奥斯汀在其《如何以言行事》的经典讲演中开创了对“施行式”话语的分析。施行式话语的特征在于：说出一句话，本身就是做一件事。当证婚人说“我宣布你们为夫妻”，他说这句话本身就是在实施结婚这个行为；他不是在描述一个预先存在的婚姻状态，而是在创造一个此前不存在的状态。奥斯汀的这套理论，对于理解预测的本体论地位有直接的启发——因为预测也在某种意义上“做”事情：它不仅仅是在描述一个尚不存在的未来，它的宣告行动本身就介入了那个未来的生成过程。

但必须立即指出预测作为一种施行式，与奥斯汀分析过的典型施行式（宣判、祝福、命名、承诺）之间的一个关键差异。在奥斯汀的分析框架中，一个施行式话语是否“得行”（felicitous），取决于它是否满足某些“适当性条件”：必须有合适的说话者、合适的语境、合适的程序。如果这些条件不具备，施行式就“失效”——但失效的方式是明确的：比如一个没有被授权的普通人对一对情侣说“我宣布你们为夫妻”，这句话的施行力就是零，婚姻并未因此成立。预测则完全不同。预测的话语在宣告之后，并不存在一个像“婚姻是否成立”那样可以明确判定的施行结果。它的“施行力”是弥漫性的：它渗透进它所处的社会场域，微妙地、不可逆地改写着人们的预期、行为和对未来的信任结构。更关键的是，预测作为施行式的“失效”方式本身就是预惑的一部分。它在“适当性条件”上

存在一种系统性的无法满足：一个预测要在奥斯汀的意义上“得体”，需要满足“所言与所指之间独立可验证”这一条件——但正如我们反复论证的那样，在时间性自反的预测中，这个条件在本体论上就无法满足。因此预惑揭示的是：预测是一种“结构性失当”的施行式——它不是在某些偶然情况下失效，而是它的构成条件本身就排斥了那种能判定其成败的干净外部基准。奥斯汀给了我们“话语如何做事”的分析工具，但他没有预见到这样一种话语：它每做一次事，就涂抹一点那个可以用来判断它做得好不好的标尺本身。预惑所命名的，正是这种特有的困境。

1.3 判别：预惑切开了什么?

现在，我们可以用一个清晰的判别表，来显式地摆出预惑与上述四个理论传统之间的分离线。

理论传统 焦点 在什么情况下停住? 预惑继续推进的方向

默顿：自我实现预言 社会因果机制：假定义→行为改变→定义成真 当一个预测不能被干净地追溯到一个“假定义”时（即预测的真假本身就取决于预测的介入） 追问预测者的本体论处境：不是“这个预测是否成真”，而是“任何预测——无论事后被判定为真还是假——都无法与它所预测的那个已在生成中被它改变的未来相分离”

麦肯齐：操演性 模型如何塑造市场：描述→采纳→市场改变→模型变准 当操演性完成了对“模型改变世界”的描述之后，不追问“模型变准”这一判断本身是否已被掏空 追问“准确”一词在操演性闭环中的本体论状态：“塑造”与“描述”这两个动作的界线在本体论上就是模糊的——模型越是“准确”，就越可能是因为它已经重塑了被它度量的那个世界

哈金：循环效应 分类如何制造被分类者：专家分类→被分类者知晓→行为改变→分类被“验证” 当分类与行为的循环闭合后，不特地区分“预测未来”与“描述现在”这两种言语行为在本体论上的差异 追问预测作为“对尚未发生之事的断言”的特有困境：事后用以验证预测的数据（如“再犯被捕率”）本身就是预测行为参与制造的——那个可资比对的、无预测发生的未来，永远缺席

奥斯汀：施行式话语 话语如何做事：说出一句话=实施一个行为；需要适当性条件 未处理这样一种特殊的施行式：其适当性条件在本体论上就无法满足（因为验证需要独立于话语的参照系，而该参照系不存在） 命名“结构性失当”的施行式：预测每一次做事，都涂抹一点判断它做得好不好的标尺本身——这不是偶然的失效，而是预测行为构成条件的一部分

这个判别表的要点在于：它展示的不是“预惑比上述概念更全面”（那只是把预惑变成了一个更大的分类格子），而是预惑在上述概念各自的分析停住的地方，切出了一个它们够不到的新辨别面。那个辨别面是：时间方向上的自我取消。 默顿、麦肯齐、哈金、奥斯汀各自以“描述现在”或“追溯过去”的认知姿态去分析“预测-对象”之间的反馈循环——他们的分析语言本身就是事后的、外在的。预惑则在预测行为发生的那个内在时间点上追问：当我的断言还在路上、未来还在生成、而我的断言本身就是这场生成的一部分时，我何以可能知道我正在做什么？这不是一个关于预测之社会效应的经验问题，而是一个关于预测行为之构成条件的本体论问题。

还有必要回应一个更根本的质疑：预惑是否只是“表征困境”在预测领域的特例？有一种反驳可能指出：所有语言和符号行为都无法穷尽其所指涉之物的全部真实性，这是表征的普遍困境。如果预

惑只是这种普遍困境在“预测”这一个特定行为上的显现，那么它就没有独立的分析价值——它只不过是“表征的本体论困境”的一个注脚。

对这一反驳，预惑的回应是：预测与一般表征之间有一个决定性的差别，这个差别使得预测的困境具有不可约简的特异性。一般表征（如“这张桌子是棕色的”）是对既存对象的描述——它的对象已经存在，且在被描述前后保持着某种本体论上的持续性。预测则不同：它的对象在描述发生的时刻尚未存在。预测是“关于一个尚未到来之物的断言”，而这一断言本身又参与构成那个它声称要描述的东西的到来方式。一般的表征困境是：语言无法完美地捕捉世界。预测的困境是：语言正在参与制造那个它声称要捕捉的世界，而它永远无法从制造中抽身出来，看一眼那个未被制造的版本。这不是程度差异，而是种类差异。因此，预惑不是“表征困境的预测版”，它是一种只有预测行为才会面对的特有困境。

1.4 预惑的边界：它在哪里生效?

在锚定了预惑不可还原的理论位置之后，必须立即补充一个理论上极为关键的边界条件：预惑并非在所有预测行为中都必然发生。 它的生效有一个严格的前提：预测所嵌入的那个系统必须是“非封闭的”——也就是说，系统中存在能够接收预测信息、并据之改变自身行为的行动者。在封闭系统中（如天体力学预测下一次日食的精确时间），预测的话语不会进入它所描述的那个系统并改变系统的运行轨迹：行星不在乎人类是否预测了它的轨道。在开放的社会系统中（如经济、司法、教育、医疗），预测的话语本身就是系统运行的一部分：央行行长的通胀预测会影响消费者和企业的行为，法官的风险评估会影响被告的未来轨迹，教师的学业预测会影响学生的自我认知。

这一边界的后果是双重的。一方面，它承认了预惑不是“一切预测”的本体论必然——它是在特定土壤（有自反性的社会系统）中生长出来的生成性矛盾。这似乎削弱了摘要中“预测行为在本体论上就必然陷入”这一强主张的字面强度。但另一方面，恰恰是这一限定让预惑变得更锋利，而不是更模糊。因为那些最关乎人的命运、最被寄予厚望以指导重大决策的预测——经济预测、犯罪风险评估、医学预后、教育分流——全都在预惑的生效域内。预惑所命名的，正是这些预测实践在构成自身时所必然卷入、却又被其技术话语系统性地排除在外的那个困惑。它不是关于某些极特殊、极反常的预测的哲学奇谈，而是关于我们这个时代最日常、最制度化、最具权力效应的那类预测的构成性诊断。

二、预惑在制度中的命运：驱离及其代价

理论锻造完成之后，我们需要将预惑推进到经验层面。如果预惑是预测行为在开放的社会系统中不可消解的生成性矛盾，那么在那些高度依赖预测的现代制度中——银行、保险公司、央行、监管机构——它是如何被处理的？本章的核心判断是：预惑作为“必须做出预测”与“无法承受预惑”之间的生成性矛盾，在当代制度中逼迫出一种特定的组织化程序——不是某个主体的主动“逃避”，而是一整套技术-官僚装置被结算出来，用以遮蔽、分散和转译预惑，使其在制度的日常运转中不被感知。我们将以2008年金融危机中风险价值模型（VaR）的生成、运作与崩溃为深度个案，追踪这一程序的完整历程。

2.1 VaR如何从一个工具变成一个驱离装置

要理解VaR为何被发明、为何在二十世纪九十年代被如此迅速而广泛地采纳，我们必须回到它被需求的那个历史时刻。在此之前，金融机构的风险管理主要依赖交易员的个人经验、头寸限额和会计账目。这些都带着浓重的“人”的痕迹——它们依赖于不可量化的直觉判断，因而也无法被标准化、比较化，无法在监管者面前被“解释”。这种状态，从制度运作的角度看，恰恰是一个预惑被全量承受的时代：一个风险经理在做决策时，不得不直面“我不知道明天市场会怎样，我不知道我的判断本身会对市场产生什么影响，但我必须现在做决定”的全部不安。这种不安不是偶然的心理波动，而是预测行为之构成条件在日常实践中的直接显现。

1994年，JP摩根推出RiskMetrics系统，将VaR——在一个给定置信水平下、给定时间窗口内可能发生的最大损失——标准化为一个可以每日报告的数字。这个数字具有一系列制度性的魔力：它把一个机构面临的所有市场风险，压缩进了一个单一的、可比较的、可答辩的美元数字。一个风险经理不需要去理解他所面对的未来有多么不确定，他只需要说出：“我们有95%的信心，明天不会亏超过五千万。”

让我们仔细审视这个数字的本体论地位。VaR说：“我有95%的信心，我们在未来一天内的最大损失不会超过X百万美元。”这个断言毫不掩饰地将“基于历史价格波动模式的、在正常市场条件下的统计推断”当作“对明天实际损失的预测”来使用。在严格的统计意义上，VaR没有做错任何事——它老老实实地报告了一个条件概率。但在制度的实际使用中，它被当作远比一个条件概率更重的东西来使用：它被当作一个可以据此设定风险限额、决定资本充足率、向监管者证明自身稳健性的“事实论断”。正是在这个从“统计工具”到“决策依据”的滑动中，预惑被有效地驱离了。因为预惑要求预测者承受一种存在性的不安：“我此刻的断言和那已经在路上的未来之间，真实的关系是什么？我的断言在多大程度上只是在用今天的数字去框住明天的未知？这个‘95%的置信水平’——它所信的是什么？”VaR则把这种不安收束到了一个简单的数字里。看着这个数字，风险经理可以不去追问概念、欲望、合法性的纠缠——他只做一件事：设定风险限额，超出即平仓。

正是在这个意义上，VaR的采纳不能被仅仅理解为一个技术选择。它是一场发生于制度层面的对预惑的集体驱离程序被逐步建立并稳定的过程。这里的驱动力不是某个人的恶意或愚蠢，而是“不得不做出预测”与“无法承受预惑”这一矛盾本身。在一个高度复杂的金融机构中，每天必须做出大量涉及未来的决策——持有或出售、发放或拒绝贷款、买入或卖出衍生品。如果每一个决策者都全然承受预惑的全部重量——追问自己的概念框架是否参与了现实的构造，审视自己的欲望如何改变了预测对象，质疑自己的认知合法性是否建立在流沙上——那么整个机构的决策速度将趋于零。预惑的无法承受性与快速决策的制度需求之间的张力，逼迫出了一种特定的组织形态：将预惑透过一整套相互咬合的技术-官僚装置进行遮蔽、分散与不可议题化。这套装置的完整版图不仅包括VaR模型本身，还包括：将欲望所在的部门（交易前台）与预测所在的部门（风险管控）做组织架构上的物理隔离——欲望和认知在组织图上被放在了不同的框框里，仿佛它们在行动者头脑中的共谋也因此被切断；将合法性的焦虑委托给外部的评级机构和监管者——不是自己在承担“我们不知道这个模型是否真的准确”的困惑，而是把这个困惑转译为“评级机构认可了我们的模型，监管者批准了我们的内部模型法”；将不可量化的直觉判断转变为可被审计、可被回溯、可被向上级演示的定量数字——从此，风险管理的全部工作可以被归档为一张写满了VaR数值的电子表格。这套装置的全部功能，不是“让预测更准”，而是“让预惑在制度的日常运转中不被感知”。

2.2 驱离的自我加固：精密化如何加深预惑

驱离预惑的制度装置并非一劳永逸。恰恰相反，它有一种内在的自我加固倾向——它必然向着更深的精密化方向推进，而精密化将累积更大的未来反噬能量。

让我们仔细追踪这一过程。VaR在实际使用中反复暴露局限：它无法捕捉尾部风险——即那些发生概率极小但一旦发生就足以致命的损失；它假设资产之间的相关性在危机中保持稳定，而实际上在系统性压力来临之时，所有资产的相关性都会骤然升高；它忽略流动性在市场恐慌时刻的突然蒸发，假设交易总能在模型设定的时间窗口内以市场价格完成。每一次模型失灵被暴露——如1998年长期资本管理公司几乎拖垮全球金融体系的事件——标准的回应都是技术升级：更复杂的尾部风险模型、引入压力测试作为VaR的补充、增加蒙特卡洛模拟的路径数量以使统计推断更稳健。这些升级无一例外地朝着同一个方向：向所有人——也向模型使用者自己——证明，“我们依然可以用现在的结构去框住未来的溢出”。一个使用了上千个参数、经过数百万次模拟的模型，会制造出更深层的幻觉：未来已经被我们精细地理解了；那些曾经渗透在直觉判断中的不安，现在可以被一个更复杂的算法所吸收。预惑在这里被压进了更深的底层——它不再表现为一种可被风险经理感知的不安，而是被转译为一项纯技术性的“模型精进议程”：模型还需要更多的数据、更复杂的算法、更细颗粒度的压力测试场景。

这就是驱离预惑的制度装置所特有的一种自我加固循环：预惑作为“必须预测又无法承受预惑”的矛盾，逼迫制度生成了第一代驱离装置（VaR + 组织隔离 + 外部评级） → 驱离装置在压力事件中被暴露为“并未真正消除预惑”，其合法性遭受震荡 → 面对这一震荡，制度的回应不是承认预惑的不可消解，而是升级驱离装置的技术精度（更复杂的模型、更细的监管标准、更严格的压力测试） → 升级后的装置以更大的技术自信，将预惑压入更不可被感知的深层 → 被压抑的预惑以更不可预测的形态和更大的能量继续累积。这不是在说制度“有意识地”在做这件事。制度没有意识。这是一个被矛盾逼迫出的、自我执行的程序——它不需要任何一个共谋者，它只需要每一个行动者在自己的位置上做“最专业的事”：风险分析师改进算法，监管者提高资本要求，评级机构更新方法论，交易员遵守风险限额。每个人都只是在做自己被制度要求做的事，而所有这些事加在一起，就构成了对预惑的更深、更不可见的驱离。

2.3 反噬：当驱离的根基瓦解时

2008年9月15日，雷曼兄弟申请破产。在此后的数周内发生的事情，不只是一个金融机构的倒闭，而是一整套认知体制的瞬间崩溃。

在此前的二十年中，全球金融体系建立了一套精细的、多层级的预测合法性机制。银行内部的风险模型度量风险，信用评级机构给抵押债务证券贴上AAA标签，监管者认可内部模型法的资本充足率计算，投资者信任评级和公开披露。这一整套机制共同完成了在常态下看起来坚固无比的事：把预惑挡在制度能够运转的门槛之外。它让人们可以不用每一刻都问自己——“我们到底在买什么？”“这个AAA标签到底意味着什么？”“我们的模型假设明天市场还有流动性，万一没有呢？”——而仅凭那些被制度认证过的数字，就能做出万亿级的交易决策。

雷曼破产击碎了这个循环。当一家被认为“太大不能倒”的机构真的倒下时，所有那些曾经被技术装置所遮蔽、被组织架构所分散、被转译为“技术精度不足”的预惑，在瞬间决堤。此前，一个投

资者可以因为某抵押债务证券是AAA级而买入它，不需要承受“这个评级本身可能在改变它所评级的东西”的困惑——评级机构已经用一整套复杂方法论为这个标签背书了。现在他发现：评级机构也在用基于历史数据拟合的模型，而历史数据中从未有过全国性房价下跌的场景——模型的“自信”建立在它从未见过真正的风暴的基础上。更糟的是，一旦所有人都开始意识到这一点——一旦预惑从一个被制度精细管理着的、不可见的背景噪音，变成了一个无人能够再声称“我不知道”的公共真相——整个信贷市场在一夜之间冻结了。不是因为资产突然变得不值钱，而是因为已经没有人再相信任何一个预测了。风险模型的数字、评级机构的评级、监管部门的压力测试结果、中央银行的通胀预测——所有这些在危机之前支撑着万亿级市场运转的认知基础设施，在同一时刻被揭掉了合法性。预惑从一个被制度装置驱离的无形困境，变成了一个震耳欲聋的、让任何决策都无法进行的公共真相。

这才是金融危机在认知层面最深刻的一个教训：不是“模型错了”，而是“驱离预惑的整套制度装置崩塌了”。即使在危机之后，改革路径也仍然在驱离的延长线上——巴塞尔协议III要求更多的资本、更精细的压力测试、更严格的风险建模标准。所有这些改革都没有触及预惑的本体论核心；它们只是在加固那套被设计来驱离它的制度机器——等待下一次积累与崩塌的循环。

三、预惑的技术重塑：当算法接过了困惑

金融模型的案例，呈现了预惑在一种特定制度形态下如何被驱离并最终反噬。当我们进入当下的“算法预测”时代，预惑遭遇的改写更为深刻。它不再只是被制度的官僚墙和模型数字所遮蔽，而是被技术性地转译——从一种存在于行动者与未来之间、可被反思的时间性困境，被转译为一个存在于数据-模型-输出闭环之中的、纯技术性的“分类误差问题”。这一转译的机制必须被具体地展示，而不是被笼统地断言。

本节以COMPAS——美国司法系统中被广泛使用的预测性警务与风险评估算法——为核心案例，追踪其从部署、争议到公共辩论的完整历程。COMPAS全称为“替代性制裁的矫正罪犯管理画像系统”，由Northpointe公司开发，用于评估刑事被告的再犯风险。它输入一百余项因子（包括被告的年龄、性别、犯罪史，以及一份由被告本人填答的问卷），输出一个10分制的再犯风险评分。法官在做出保释、量刑和假释决定时，会看到这个分数。

3.1 预惑如何被转译：技术-制度装置的接力

2016年，ProPublica的调查记者朱莉娅·安格温与其同事发表了一篇里程碑式的调查报告。他们获取了佛罗里达州布劳沃德县使用COMPAS评分的七千余个真实案件记录，并将风险评分与被告在随后两年内是否再犯的实际数据进行比对。其核心发现是：COMPAS的预测在被标记为“高风险”的被告中，黑人与白人之间存在系统性的误判差异。在被标记为高风险但并未再犯的被告中，黑人被告的比例是白人被告的近两倍；相反，在被标记为低风险但实际上再犯的被告中，白人被告的比例更高。

这一发现当然立即引爆了关于算法偏见的争论。但本文在此处要追问的不是“算法是否公平”——这个问题已经被大量文献充分讨论。我们要追问的是：在COMPAS从部署到争议的整个过程中，预惑被如何处置了？答案不能仅仅是“被遮蔽了”——这个说法太过笼统，它把那种独特的、技术-制度协同运作的转译过程模糊掉了。我们需要具体地拆解：预惑从一个行动者与未来之间的时间性

困境，被转变为一个数据-模型-输出闭环内的技术问题——这场转译究竟是被哪些技术特征和制度安排共同完成的。

第一步：量化输出将“未来”替换为“分数”。 COMPAS输出的不是一个需要人类法官去诠释、掂量、质询的风险叙事（如传统量刑前调查报告中的长文评估），而是一个1到10的数字。这个数字具有一系列认知效应：它把“这个人未来是否会犯罪”这个充满了时间性自反的不确定问题，转换为“这个人在一个10分制量表上的位置”。面对一个数字，法官被卸下了许多追问的负担：他不需要去想象被告的未来的多种可能性，不需要去掂量“我所看到的犯罪史证据”与“这个人实际上可能改变的可能性”之间的关系，不需要去质询“我此刻依据这个数字所做的判决”会如何反过来影响被告的未来。他只需要将那个数字与量刑指南上的分档对位。预惑中那个“因果账该如何计算”的不可消解的不安，在此被数字的表观确定性所遮蔽。

第二步：技术黑箱将因果链条变为不可审查之物。 COMPAS的算法权重是专有商业机密，Northpointe公司拒绝公开其计算公式。这意味着，不仅被告本人无法知道究竟是哪几个因素——是他之前的犯罪次数，还是他的年龄，还是他自陈的“同辈犯罪程度”——导致他被评为7分而非3分；法官同样不知道。当一个预测的生成过程成为一个黑箱时，任何想要追问“这个预测本身在多大程度上是它所预测的那个结果的制造者之一”的尝试，都被技术性地阻断了。你无法追问一部你不被允许打开内部构造的机器：“你对你的预测对象的介入，如何改变了你声称只做评估的那个未来？”这种追问在COMPAS的制度安排中甚至没有可以进入的端口。

第三步：专业责任在打分者与用分者之间弥散。 在COMPAS的制度安排中，评分由Northpointe公司开发的算法自动生成；法院是算法评分的使用者，但不是算法的开发者，也不是算法准确性的验证者。当一个被告提出“这个分数对我不公平”，法官可以合理地说：“我只是依据风险评估工具得出的客观分数进行量刑。”而算法开发者可以说：“我们的算法在统计上是准确的——它的预测准确率经过了测试。”预惑——那个“预测正在改变它所预测的未来”的责任——在这套制度安排中没有可以被归置的主体。算法的开发者负责模型精度，法官负责依法裁决，但谁负责那个“我的预测一旦做出就不可逆转地改变了这个被告的未来轨迹”的本体论事实？没有人。这不是有人渎职，而是这套技术-制度装置的结构中就没有为这个问题预留位置。

第四步：一切异议被转译为“分类误差问题”。 当ProPublica调查报告揭示COMPAS存在种族间误判差异后，公共辩论迅速集中到了“算法偏见”和“模型公平性”上。Northpointe公司的辩护也完全落在这个层面：它发布了一份技术反驳，指出ProPublica的分析方法有误，并强调COMPAS对黑人和白人的“预测准确率”实际上是相同的——即在黑人和白人被告中，被标记为高风险的被告最终再犯的比例是相近的。论辩双方使用相同的分析语言：准确率、误判率、假阳性、假阴性。在这场论辩中，预惑完全消失了。不是因为有人刻意压制它，而是因为这场辩论的制度框架——算法公平性的技术争议——甚至没有一个让预惑可以进入的语言端口。“这个预测是否正在参与制造它所预测的再犯结果”这个问题，无法被翻译成公正性的技术指标，无法被放进假阳性和假阴性之间的权衡公式。因此，这个问题就从公共理性的视野中被驱逐了。

通过这四个技术-制度装置的接力——评分被量化为一个数字、算法被封装为黑箱、专业责任在机构间弥散、所有异议被转译为技术性的分类误差问题——预惑完成了从“时间性困境”到“分类误差

问题”的完整转译。它不再表现为一种可以被法官、被告或公众感知的存在性困惑；它换上了一件纯技术性的外衣，进入了一个唯有数据科学家才能参与辩论的场域。

3.2 可解释性与预惑的不可通约

面对算法偏见的公众压力，一支庞大的技术伦理工业迅速崛起，其核心旗帜是“可解释AI”。在COMPAS的争议中，主流回应是呼吁“算法透明”：让风险评估模型的公式公之于众，让被告知晓哪些因子影响了对其风险评分的计算，开发可用于解读模型决策的局部解释工具。这一套话语迅速被立法者、技术公司和学术研究者采纳，成为应对算法争议的标配答案。

现在，我们必须严肃地追问一个更深层的问题：提高模型可解释性，是否能够触及预惑？ 本文的论证是：不能。原因不在技术，而在本体论。

可解释性提供的“解释”，在任何现实操作的、非封闭的系统中，都是事后构建的、服务于人类认知的局部叙事。即使COMPAS的公式完全公开，即使每一个输入因子的权重系数都被公之于众——前科次数、年龄、教育程度、药品使用史、被告对其同辈犯罪行为的自述打分，每个变量对最终风险评分的贡献百分比被精确地分解并呈现——即使我们确切地知道模型在计算“本被告有87%的概率再犯”时依据的是什么数学运算，这个解释能告诉我们什么？它能告诉我们：在这些历史数据中，具有类似特征的人群有更高的再犯被捕率。但它不能告诉我们：这个“被捕率”在多大程度上是前一轮、再前一论的预测性警务实践的产物——在过去几年中，由于警力更密集地部署在被标记为“高风险”的社区，那些社区中的人即使犯了和被标记为“低风险”社区中的人完全相同的罪行，他们被捕的概率也更成倍地更大。它不能告诉我们：如果我们不再使用这套算法、转而投入同等资源去社区矫正和社会支持，这个被告的再犯概率会是多少。它也不能告诉我们：眼前这个活生生的被告——他此刻正站在法官面前，刚刚听到自己被一个算法评估为“高风险”，他的家人、他的雇主、他的假释官也都听到了这个数字——他走出法庭的那一刻，他的未来已经被这个数字改变了多少，而那个未被改变的未来，是一张永远无法冲洗的底片。

可解释性所做的工作，是把模型内部的数学运算翻译成人类可以理解的故事。而预惑指向的，是翻译永远够不到的那个东西：预测行为与它所预测的未来之间那段被污染的、不可分解的因果链。 可解释性可以告诉你“模型为什么这样预测”，但它不能告诉你“这个预测本身在制造着什么”。前者是技术可解的问题，后者是本体论上无法被算法化的问题。把可解释性当作对预惑的回应，是在把本体论困境误诊为技术不透明——是在给预惑穿上一件人类能理解的衣服，然后宣布“困惑已经消除”。

3.3 被驱离者的回归：当预惑在法庭上爆发

然而，预惑不可能被完全消除。即使在黑箱最严密的封口之中，它也会在某些裂缝中爆发。在COMPAS的争议中，预惑的爆发点不是技术性的，而是主体性的：一个人，面对一个决定自己命运的算法分数，发出了不可被消解的质问。

2016年，威斯康辛州最高法院审理了State v. Loomis案。被告埃里克·卢米斯因驾驶一辆用于枪击案的车辆而被捕，在辩诉交易中认罪。随后，法院使用COMPAS对卢米斯进行了风险评估，得出的分数标记他为“高再犯风险”。法官在量刑时引用这一分数，判处卢米斯六年监禁和五年延长监管。卢

米斯提起上诉，其核心论据之一是：COMPAS的算法机制是专有秘密，他被剥夺了检验对他的指控性评估的权利——他无法质疑那个将他标记为“高风险”的机器；他不被允许知道，那个决定了他刑期长短的分数究竟是如何被算出来的。

让我们细读卢米斯上诉书的潜在逻辑。在法律的技术话语之下，他的质问其实正是预惑在人被抛入这场算法预测时所爆发出的原生形态：“这个说我会犯罪的数字，它是什么？它凭什么决定我的未来？我如何在它的面前为我尚未做出的行为辩护？如果这个数字改变了我被对待的方式——更多的监控、更少的假释机会、更重的刑罚——那么当它‘预测成功’时，它是在报告一个关于我的事实，还是在报告一个它自己参与制造的结果？”法庭的裁决——驳回上诉、维持原判——在技术上肯定了算法辅助量刑的合宪性，但在本体论上，它标志着预惑在当代司法中的一种特定命运：它被制度封存在了“风险评估工具的统计准确性已经经过验证”这一技术性事实的背后。卢米斯输了官司，但他的质问并没有消失。那个质问——那个“这个数字，它是什么？”的呐喊——正是预惑从一个被技术-制度装置转译为“分类误差问题”的东西，重新爆发为一个不可压制的主体性问题的时刻。

卢米斯的质问不是孤例。在预测性警务和风险评估算法日益渗透进司法系统的过程中，每一次算法失灵被曝光——一个被COMPAS标记为“低风险”的人在保释期间犯下了暴力罪行；一个被“高风险”标签追踪了多年的年轻人，最终发现他在整个监控期间从未有过任何违法行为——这样的时刻，预惑都以一种公众愤怒和认知震惊的方式复苏。人们突然意识到：那个我们一直在说的“准确”——COMPAS的开发者声称其模型达到了某种百分比的“预测准确率”——到底是什么意思？是谁的准确？是以谁的被改变的未来为代价的准确？如果预测的准确率本身就依赖于一个已经在前一轮预测中被系统性扭曲了的“再犯”数据集——在这个数据集中，被标记为高风险的黑人被更频繁地逮捕，因此也更可能出现在“再犯”记录中——那么，“准确率”这个概念本身，还能在什么意义上被信任？这不是对技术细节的好奇，而是预惑的公共爆发：整个社会在那一瞬间发现，它已经把自己对未来的判断力，外包给了一种它根本不知道其本体论资格的机器——而当它试图事后追问这台机器的资格时，唯一的回应是：“它的准确率是有数据支持的。”

结论：在预惑中行动

本文的论证轮廓收束于此。“预惑”，不是一项可以被解决的“问题”。它是预测行为在时间性自反结构中不可消解的生成性矛盾。只要人类还在开放的社会系统中做出预测——无论是用龟甲上的裂纹，还是千亿参数的深度神经网络——预惑就与预测同在。它不被技术升级所消除，不被制度设计所消解；每一次对它的驱离，都以更大的尺度积累了它将重新闯入公共视野时的能量。

这一分析视角的意涵可以往三个方向上打开。

首先，它要求我们彻底反思预测在社会决策中被赋予的“事实提供者”角色。在法律、金融、教育、公共卫生等大量制度化预测实践中，预测数字被当作一种可以与“客观事实”进行比较的论断。但一旦我们承认预惑的结构性存在，就必须同时承认：预测从来不是“关于”独立于预测而存在的未来事实的中性报告。它是一场介入，是在未来正在生成的过程中投入了一个改变生成轨迹的力量。将预测当作“事实提供者”，就是用一种错误的分类，掩盖了它作为“介入者”的实际存在方式。这不是一个技术修正的问题——不是说以后的风险评估报告应该在末尾加上一句“本预测可能扰动被预测

现实”的免责声明。免责声明的本质仍然是发现学的：把预惑当作一个外部追加的风险来处理，而不是承认它就是预测在构成自身时所同时制造的那个不可消解的盲区。

其次，预惑打开了对“理性行动”这一概念的重新理解。传统的决策理论将“理性行动”建立在预测之上：收集所有可获信息，预测各种行动方案的后果，选择带来最优期望后果的那一个。但预惑告诉我们，在最需要预测的那些决策中——也就是说，那些后果不可逆、且决策本身就会改变后果生成条件的决策——预测的“最优性”建立在被它自己涂抹掉的那根标尺上。那么，在预惑的侵彻下，决策伦理在实践中可能被重新结算为怎样的形态？一个方向是：从“选对”转向“带着回应责任介入”。当决策者承认自己不占据一个“可以干净地看清楚后果”的外部位置时，行动的重心就不再是计算后果的精度，而是对所介入的共同体与未来人所负的持续回应责任——不是“如果预测错了就追责”的事后追责，而是一种嵌入干预过程之中的、对“我的行动正在制造它将要验证的那个未来”这一事实的警觉。这不是在宣告一种“真正的理性”，而是在描述一种在预惑被承认的压力下，决策者的自我理解可能被逼出的走向。

第三，预惑不是一项关于预测的社会病理诊断，而是一种清醒的社会认知的构成条件。一个社会如果成功地用制度与技术完全驱离了预惑——让每一次预测都看起来只是“科学”“客观”“精确”的中性操作——那么这个社会就处在对自身认知条件的最深无知之中。在那种无知中，未来被当作一个可以无限预纳和精算的对象，决策者被卸下了“我可能正在参与制造我声称只做预测的那个未来”的全部负担。而当这种负担被完全卸载时，决策的反噬就只是时间问题——它会在某个压力时刻，以人们对整个知识体系合法性信心崩塌的形式集中爆发。

因此，对于当代社会中密集依赖预测的治理、商业、司法和日常决策而言，预惑所要求的不是“解决方案”，而是一种制度层面的公开争辩空间。当一个公共机构部署一套预测系统时，它目前通常只做两件事：评估模型的技术精度，并根据预测结果采取行动。预惑要求在这二者之间，插入一个被制度保护的第三环节：一种对未来预测的本体论地位的公开审问机制。接受这一预测的人——无论是被告、学生、贷款申请人，还是更广义的被预测所影响的社区——是否有权利知道，并且有制度空间去辩论：这个预测改变的，可能正是它声称只做客观评估的命运本身？这不是关于“算法透明度”的又一个委婉说法。算法的技术细节公开解决的是信息不对称，但它不解决预惑——因为它不能回答“这个预测本身在制造着什么”。预惑的公共容纳，要求的是另一类实践：一套独立于模型开发者与部署者的、有权质询预测的本体论资格的审视机制；一种在公共理性中为“这个预测正在制造什么”保留持续争论权利的制度安排；以及一份永远不可被剥夺的、个体和群体选择不进入某种预测体制的权利——不是因为反对“科学”，而是因为选择自己承担预惑的全部重量，而不是把它外包给一个无法为其行动负责的机器。

最后，本文对预惑之本体论地位的核心主张，在以下条件可被质疑乃至证伪：如果在任何一个具备时间性自反性的非封闭系统中，能够持续地做出对未来具体事态的精准断言，并且该预测的公开宣告、被行为主体广泛采纳后，其对被预测现实的扰动可以被独立于预测者及预测模型的方式测量、并严格地与预测值分解开来——从而确凿地分离“预测所抵达的那个未来”与“预测未发生时本应发生的那个未来”——亦即预测行为在本体论上并未陷入时间方向上的自我取消。如能系统性实现这一分离，则本文关于预惑为开放社会系统中预测之构成性困境这一判断，在该领域失效。这一条件的严苛性本身，恰恰说明了预惑在可预见的未来中，仍将是所有严肃的预测实践无法回避的构成性困境。

注释

[1] 本文对默顿（Merton, 1948）的对话仅限于自我实现预言的经典表述，不涉及默顿后期关于中层理论与社会结构分析的全部体系。

[2] 麦肯齐（MacKenzie，2006）的操演性理论是本文1.2节的核心对话对象之一。本文对麦肯齐理论的解读仅限于其关于金融模型操演性的论证，不涉及麦肯齐关于核导弹制导精度之社会建构的其他著作。有关操演性在经济社会学中的更广泛应用，可参见Callon (1998) 与MacKenzie, Muniesa & Siu (2007) 所编文集。

[3] 哈金（Hacking, 1995, 2006）的循环效应与动态唯名论是本文1.2节的另一核心对话对象。本文仅取其关于“分类如何制造被分类者”的论证，不涉及哈金在科学哲学与概率论史的广泛工作。

[4] 奥斯汀（Austin，1962）的言语行为理论对本文1.2节的理论锻造提供了关键参照，但本文不进入塞尔等人对言语行为理论的后续修正与发展。

[5] （材料说明）本文第二、三章中的案例——2008年金融危机中VaR模型的制度实践及COMPAS算法争议——均基于公开历史记录、媒体报道、学术评论及司法文件的重构与分析。为突出理论逻辑，相关过程细节经过简化和典型化处理；文中未使用任何一手访谈或未经公开的实证数据，所有事例均作为思想阐发性质的例示，不应等同于经过同行评议的经验验证。第二章中关于JP Morgan内部决策过程的描述，系基于理论逻辑的重构而非基于档案的历史证明。第三章中关于COMPAS算法转译机制的分析，系基于ProPublica调查报告、Northpointe公司公开技术反驳及Loomis案法庭文件的概念图解，Northpointe公司内部的设计决策过程不在本文考察范围内。

附论·最近邻切割：预惑不是自我实现的预言、不是反身性、不是操演性

「预惑」这个新造词，站在三个已经很有名的邻居的门口。若不划清界限，它随时会被读成它们的同义反复。

第一个邻居是默顿的「自我实现的预言」（self-fulfilling prophecy）。默顿说：一个原本为假的预测，因为人们相信它并据此行动，最终变成了真——银行挤兑的经典案例。本文的分离线：自我实现的预言是一个关于结果的命题（预测改变了它所预测的世界，使结果与预测吻合），它的落点在预测与世界的因果闭环；「预惑」是一个关于预测者处境的命题（预测行为在本体论上必然把预测者卷入自己所预测的时间之中，使他无法站在一个外部时点观测未来）。默顿的预言者仍然站在世界之外发出预言，只是预言反弹回来改变了世界；预惑的预测者根本没有「世界之外」可站——他做预测的这个动作本身已经是被预测的那个未来的一部分。可裁决的分离：自我实现的预言可以靠「保密预测、不让被预测者知道」来消解（不公开就不反弹）；预惑不能靠保密消解，因为困惑的根不在预测是否被公开，而在预测者与被预测时间的本体论纠缠——哪怕预测绝对保密，预测者调配资源、布置预案的动作也已经改写了他所要预测的那个未来的初始条件。

第二个邻居是社会学的「反身性」（reflexivity，尤其索罗斯的市场反身性与吉登斯的制度反身性）。反身性说认识与对象互相构成、认识反过来塑造对象。「预惑」看起来就是反身性在预测领域的应用。本文必须切开：反身性是一个双向因果的结构描述（认识→对象→认识……的环流），它是

价值中立的、可正可负的；「预惑」不是描述这个环流，而是指认这个环流对预测者构成的一种特定的、不可解的困境——预测越精确、越被采信、越被据以行动，它就越深地改写自己要预测的未来，从而越确定地使自己失准。反身性是中性机制，预惑是这机制在预测这一特定行为上逼出的自败结构。分离线可裁决：反身性框架预测「认识与对象互相塑造」，不预测方向；预惑预测一个特定方向——预测的采信度与其长期准确度在结构性转折点上负相关（越被信、越在转折前夜自信地报「一切正常」）。这个负相关是可测的，测出来正相关或无关，预惑就被证伪。

第三个邻居，最危险，是巴特勒/奥斯汀的「操演性」（performativity，及卡隆、麦肯齐把它引入经济学的「模型塑造市场」）。麦肯齐已经证明布莱克—斯科尔斯模型不是描述了期权市场、而是使市场变得符合模型。「预惑」会被直接读成操演性的老调。切开的关键：操演性讲的是表述如何生产它所描述的现实（模型使世界像模型），它的方向是从符号到现实的成功塑造；预惑讲的是这种塑造成功之后必然反噬预测自身的那一步——正因为预测操演地塑造了世界，预测者才更深地失去了预测那个已被自己改写的世界的能力。操演性到「模型造出了它的世界」为止是一个成功故事；预惑接着说「而这个成功恰恰是预测自败的机制」，它是操演性的下半场、是操演成功之后的本体论账单。可裁决线：操演性预测「被广泛采纳的模型会使市场向模型收敛」；预惑追加预测「这种收敛会在收敛最完全时最脆弱——即模型塑造得越成功，系统对模型未覆盖的扰动越无抵抗力」。这条追加线在 1998 年 LTCM、2008 年危机中可被回溯检验，也可在任何「模型高度统一的市场」里前瞻检验。

附论·可裁决预测：本体论困境也要能被经验判错

「本体论困境」这种大词最容易滑向不可证伪。本文偏要给它三条经验裁决线。

预测一（采信—失准的负相关）：一个预测系统被采信并据以行动的程度，与其在结构性转折点上的准确度负相关。操作化：测量某类预测（如央行通胀预测、系统性风险模型）的市场采纳度，与其在事后被确认为「结构性转折」的时点上的误差。本文预测采纳度越高、转折点误差越大。若测得正相关或无关，预惑被证伪。

预测二（保密不解困）：把预测严格保密（不让被预测的行动者知晓）并不能显著改善其在转折点的准确度，因为预测者自身依据预测所做的资源调配已改写初始条件。若保密预测在转折点上显著更准，则「困境源于预测者与时间的本体论纠缠」这一主张被削弱——困境就只是「公开导致的反弹」，可由保密消解，不配称本体论困境。

预测三（收敛—脆弱耦合）：模型在一个领域内被采纳得越统一（所有玩家用同一套预测框架），该领域对模型未覆盖扰动的抵抗力越低，转折来临时的崩塌越剧烈。若统一采纳的领域反而更稳健，则预惑对「操演成功即自败前夜」的刻画错误。

参考文献

[1] Merton, R. K. (1948). The self-fulfilling prophecy. The Antioch Review, 8(2), 193-210.

[2] MacKenzie, D. (2006). An Engine, Not a Camera: How Financial Models Shape Markets. Cambridge, MA: MIT Press.

[3] Giddens, A. (1990). The Consequences of Modernity. Stanford: Stanford University Press.

[4] Austin, J. L. (1962). How to Do Things with Words. Oxford: Clarendon Press.

关于本文的创新智商标注

本文在原始投稿基础上，由编辑依王德生《SDE本体论》与 SDE 创新智商评估规程做了「最近邻切割」与「可裁决预测」两节加固，意在抬高其不可还原性（I）与可证伪性（F）两维。依评分规程铁律，任何人不得为自己参与修改的文本认证分数，故本文页面所涉创新智商为**修改设计目标（135–139 区间）**，非认证分；认证分须由独立评分者盖出处另行给出。