

前言：演进社会中的制度设计与实施

张曙光 丁利

马克思有句名言：哲学家只是用不同的方式解释世界，而重要的在于改造世界。这个世界的奇妙之处的表现形式之一恰恰就是如玻尔所说：在生活的舞台上，我们既是演员，又是观众。生活在一个非严格决定论的世界里，我们既是行动者，同时又是解释者。能够使我们为世界有所改造的，除了我们的雄心，我们所倚赖的不过是我们解释世界的知识而已。在我们的知识之外，我们把决策选择交给我们骨子里的冒险或谨慎的天性以及自然那不可蠹测的偶然。

处于制度变迁过程中的中国社会，可能更真切地凸显出这样一个看似佯谬的自我相关的特征。通过对制度变迁案例的系统研究，我们有可能对我们所在的社会“窥一斑而知全豹”，从而将我们的经验、教训凝结而成的知识回报于社会和实践。本书的安排是这样的。首先，我们从元理论的视角，勾勒出关于演进社会中的制度设计与实施问题的某些重要侧面。正因为要取一个元理论的视角，我们最终得出的不是一些具体的结论，而更多地是一般原则。然后，我们对第三批案例研究作出一个简要的评述。

一、演进社会中的制度设计与实施的一个简要框架

按照哈耶克（Hayek, 1967）的看法，“社会理论的整个任务，乃在于这样一种努力，解释整个经济活动的秩序是如何实现的：这个过程运用了大量的知识，但这些知识并不是集中在任务单个人脑中的知识，而仅仅是作为不计其数的不同的个人的分立的知识而存在的”。在我们看来，把这一问题表述清楚的难度与现实里人们如何解决它的难度差相比拟。

这段话至少隐含着几个意思，一是在分工和专业化的社会里，掌握分立的知识的人是如何传递信息的，或者说是如何仅仅依靠少量的信息就协调起来了；二是在博弈论所谓的互动社会里，个人理性决策是建立在关于其他人的行为的信念的基础上的，而其他人的行为又是建立在他们关于别人行为的信念的基础上，如此无穷，这样一个自我相关的无穷回归问题是如何得到解决的颇耐人寻味（特别在存在多重均衡的情况，人们的实际博弈中采取什么均衡策略）；三是既然一个人不掌握别人所具备的与他利益相关的决策的知识，他是如何相信别人不会利用其独有的知识做出对他不利的决策。换言之，他是如何了解并确信博弈的支付函数的。

笼统地说，我们可以认为社会是通过制度自我组织起来的。近几十年来，博弈论在知识论和演进论两个发展方向上的成果使我们对均衡，或者人们的行为模式的形成有了新的认识。把博弈论与社会选择、机制设计等领域的成果结合起来，使我们有了分析制度问题的新工具，也对制度的本质有了更深刻的了解。

我们区分关于制度的两个不同层次上的理论，即规范理论和实证理论，并试图探讨它们之间的复杂关系。我们首先给出一个

关于主流的机制设计理论及其扩展的粗浅概括，然后通过讨论如何理解均衡与制度，我们尝试说明主流经济学关于制度的一般理论框架在什么意义上可以认为是对现实的合理解释；进一步，我们通过把制度演进转化为知识进步，讨论变革社会制度演进过程的动力学机制，并希望同时表明其外部表现形式的复杂性。

（一）经典机制设计理论及其扩展。

正如囚徒困境博弈所揭示的，个人理性的行为有时候会导致社会的集体低效率^①。那么超个人的力量的出现，如政府、法庭等所谓制度性组织，在规范意义上正是为了解决诸如此类问题而出现的。

经过赫尔维茨、马斯金等人的工作，制度的规范问题已经转化成这样一种经典表述：如果按照社会选择理论，对于每一种社会经济环境，我们能够确定一个满足某些价值规范的社会目标集合^②。并且我们接受人们在互动的社会中是按照博弈论所刻画的方式行为的观点，那么，机制设计（mechanism design）理论则探究如何提供一个博弈形式（game form），使得这个博弈形式下的均衡解是在社会选择目标集合里。也就是说，社会选择函数是可执行的（implementable），或者退而求其次，这种均衡解无限接近于社会选择目标集合，也就是说可以近似地（virtually）执行^③。

从社会选择理论我们知道，如果存在一个社会选择函数（或对应），那么，在任何一种情形下，针对个人偏好组合，我们都有一个或几个社会结局是我们认为合理的。但这是从一个客观看

① 杜比证明，几乎所有博弈的纳什均衡都是非帕雷托最优的。

② 例如（Sen, 1986）。

③ 参见考尔肯（Corchon, 1996）教科书性质的综述文献

察者或研究者的角度得出的结论。如果社会的设计者 (designer) 或计划者 (planner)、执行者 (executive or implementor)、仲裁者 (arbitrator)，像客观观察者一样，具备对社会的足够知识和信息，那么一个简单的、集中的强制性机制（如果是技术上可行的）就会实施任何所谓合理的社会目标。但是，真正有意义的是，设计者或执行者不可能具备这么完全的知识和信息，或者有些信息是他由于某种原因不能使用的^①，或者信息是可观察的 (observable) 但不是可确证的 (verifiable)。简言之，我们面对的基本问题是不可避免的分散决策 (decentralized decision)。这时需要探讨的是，如何使得通过社会选择函数表现的社会目标能够实现。机制设计理论把这个问题转化为，假设设计者知道社会选择函数，而非特定社会状态本身，是合理的社会目标，并且他知道社会的行为模式，即行为主体是如何做出行为选择的以及任何行为选择会导致什么样的结果，那么，我们只需要设计一个制度框架，例如，财产权利、契约法律、投票法律和投票规则以至宪法，使其影响人们的行为选择从而间接实现社会目标。

不能像一个客观观察者一样作为一个限制条件，带来至少两个问题：第一个是信息 (information) 问题。如果不考虑具体环境的特殊性，那么仅仅从保证机制正常运行的角度，应该使必要的信息得以顺利流通，而信息流通的一个基本要求是传输它的信号空间 (message space)，也就是信息得以交流的语言。在一个社会的技术约束下，我们自然希望信号空间的维度 (dimension) 越小越好。另外，博弈者的有限理性使得他所分析处理的对象应该首先是可计算的 (computable)；其次，应该是计算复杂性 (computational complexity) 程度受到限制的。第二个问题也是我

^① 例如，制度禁止的。

们研究的重点是激励 (incentive) 问题。正如梅耶森 (Myerson, 1985, 229) 所总结的: “在社会和经济事务中有两种激励约束限制了人们达到互利协议的能力。首先, 当一个人有了他人所没有的不能确证的私人信息时, 那他就不会被说服诚实地显示这个信息, 除非给予正确的激励。其次, 当一个人左右着一个秘密的决策变量而其他不能控制或监视时, 那么他就不会被指挥着选择任何特定决策或行动, 除非他被赋予激励这样做。一个社会契约或协调系统, 如果给予人们在他们的信息上撒谎或者在行动上欺骗的激励, 那它就是不可行的。一个组织必须赋予其成员正确的激励使之分享信息并一致行动。一个人不能被指望提供对己不利的证据或者付出得不到回报的努力”。

我们希望的是, 一个机制在存在种种问题时依然能够执行合理社会目标^①, 所以我们需要关注的核心问题之一是机制的稳健性。如摩尔 (Moore, 1992, 210) 指出, 很多与显示机制接近的机制都没有在实践中运用, 是因为它们缺乏稳健性。 “复杂的机制可能在现实的不完美世界里执行得很糟糕, 或者因为博弈者的理性中有缺陷, 或者是在博弈者的偏好、知识或处境 (博弈者能在机制外交流吗? 他们可以共谋吗? 他们能够影响自然状态的实现吗?) 的描述中产生错误”。好的机制应该具有好的容错能力, 即不会因为小错误而导致大的损害^②。

马斯金和摩尔研究了存在再磋商 (renegotiation) 情形的执行问题。马斯金和摩尔 (Maskin & Moore, 1999) 所指出, 执行问题的文献大多排除了再磋商的可能性。这是一个主要作为偏

① 一个合理的社会目标能够在不同机制里被执行似乎也是很有吸引力的。

② 非线性动力系统理论, 特别是混沌理论, 给我们的启发是, 这个目标不是那么容易实现的。作者将在另外关于博弈论基础的文章中专门讨论此问题。

离均衡现象而涌现的问题。当博弈者设计一个制度或契约时，他们被假设为关注于保证帕雷托最优的结局，这样一个执行机制的均衡结局在没有再磋商机会的意义上是有效的。但在偏离均衡时，结局可能远非帕雷托最优。事实上，执行理论的文献中，机制有时就通过给予偏离均衡的策略组合一个坏结局作为不鼓励偏离的手段。但是假设博弈者发现他们在一个偏离均衡的位置上，此时他们面对一个无效率结局的前景，如果有另一个他们都喜欢的可能结局他们为什么应该坚守在这儿呢？换种说法，他们为什么不应该简单地撕毁原先的合同，然后协商一个新的以实现帕雷托改善呢？

再磋商的发生，既可以理解为博弈者之间不能做出一个有约束力的承诺（commitment），也可以归结为制度执行者的某种能力不足。这就相当于原来的博弈结构没有完全描述整个博弈的场景，需要在某些节点上加上一个再磋商的子博弈。所以，在马斯金看来，完全理性的博弈者应该能够承诺不发生再磋商（Maskin & Moore, 1999）。但还有一个再磋商存在的合理理由，那就是博弈者偶尔会犯错误（特别是很多时候难以区分故意或过失，如果对过失施以严厉惩罚，似乎会促使人们为了降低风险而减少互利交易的可能性），此时允许再磋商对社会福利是一种改善。所以，在现实社会的许多法律中再磋商是被允许的。

通过对包括再磋商的博弈机制本身施加三个自然的假设，即再磋商是可预见的、有效率的和个人理性的，马斯金和摩尔（Maskin & Moore, 1999）在附加一些很自然的其他条件下证明了一个与纳什执行类似的定理，一个社会选择规则是再磋商均衡可执行的，当且仅当它满足再磋商单调性。

从其他角度，泰罗尔（Tirole, 1992）总结了存在共谋情形下的组织理论。斯尧斯特罗姆（Sjostrom, 1999）研究了博弈者

存在共谋和再磋商情形下的机制设计，而拉方特和马尔迪摩尔特 (Laffont & Martimort, 2000) 则研究了博弈者共谋和类型相关情形下的机制设计。贝里伽、考尔肯和斯尧斯特罗姆 (Baliga, Corchon & Sjostrom, 1997; Baliga & Sjostrom, 1999) 等人考虑了“互动执行”，即计划者成为博弈者的情形。通过假设计划者在博弈中的效用函数就是社会目标，他们得出了在不同行为模式（博弈的解概念）下互动执行的充分必要条件。

一个机制要有意义，其最基本的要求机制中的结局函数是可确认的，从而可以给博弈者可预见的结局。这就隐含地使得制度的实施者的作用成为必要（当然，一个最特殊的实施者是“自然”或所有博弈者）。如赫尔维茨 (Hurwicz, 1992) 所指出，“一个更严峻的，我们在实践中也都很好地明白的问题是，那些被交付实施或信息处理责任的人与那些被期望作出规则行为的人，可能会以被分别称为共谋、腐败或指派 (being co-opted) 的方式联合起来。另一个需要考虑的可能性是，那些负有实施责任的人对他们被期望实施的规则和条例冷淡视之”。赫尔维茨提出一个“博弈层次”的框架，第一个博弈的结局空间是下一个博弈的“规则的规则”，如此进行下去，伴随着博弈结局的特殊性的增加，直到最后的博弈实质性的结局（例如资源配置）才被确定。如果我们相信，第一个博弈具有无可比拟的优先性，那么，这个问题可以转化成，一个社会的所有人和一个或几个（在分权制衡的思路下）特定博弈者之间的最优机制设计。分析这个问题的合适框架是契约理论和委托—代理理论。严格来说，还需要一个机制设计的前博弈使结局函数是可确认的，从而使这个博弈成为可实施的。这相当于立宪会议或历史传统等因素所形成的使某个制度能够得到执行的更基本的制度。更确切地说法是，某个表层制度（机制）的可实施性源于更具基本性的制度，后者把前者

中的执行者变为一个博弈者，这个基础性博弈的结局决定了前者的结构（策略空间和结局函数）。最终我们可以追溯到一个自我实施（self-enforcing）的制度，这个制度是由所有博弈者的历史赋予的最基本和最稳定的知识和技术水平所决定的。这就进入关于制度演化的实证理论了。

所以，值得指出的一点是，机制设计理论的发展似乎表明，不管什么样的社会目标（社会选择规则），总能找到执行它的机制；而不管机制中对人们的行为模式是如何预设的，总能找到它能执行的所谓最低合理限度的社会目标。特别是几乎所有社会选择规则，在精炼纳什均衡行为假设下都是可执行的，或者总是可近似执行的。我们似乎处在莱布尼兹的乐观判断中了：我们所生活的世界是所有可能世界中最好的那个。

当然，机制设计理论中也有种种“不可能性”定理，说明问题不那么简单。但怎么理解看似冲突的不同命题呢？我们在后面将试图给出一个分析此戈尔迪亚斯之结的思路。

（二）作为进化过程中稳定均衡的制度。

从研究者的角度看，机制设计理论是存在性的，它只表明研究者作为一个系统外的客观观察者，可以确信存在一个机制以实现某些目标。研究者自己做了制度设计者的角色，或者说研究者提出了一个对制度实施者的建议。这样一个建议在现实中是如何得以实现的，本质上是我们在后面所讨论的知识进步的一个特殊部分。现在我们关注的是其实证意义。因为我们作为研究者，首先是要对世界提出解释。而机制设计理论的许多结论，只要进行适当的处理就可以对历史上和现实中存在的制度作出合理的解释。

问题的关键在于，我们如何理解博弈的均衡解概念。在知识

论^①背景下的博弈论中，一个博弈的均衡解，通常要求博弈者了解博弈的结构，包括博弈的参与者及其偏好和策略集合，建立在自然状态和策略组合上的结局函数等等，并且这些都是“普遍知识”（common knowledge），即每个人都知道这些，以至无穷。在这样的条件下，均衡策略包含了两方面的内容，一是每个博弈者的行动方案；二是策略代表了其他博弈者关于他的可能行动的信念。所以，均衡解作为一个自我实施的行动方案组合，也就成为博弈者所在社会的普遍信念。

显然，这个知识论前提加上其他条件（如逻辑全知）的现实性是值得深思的。严格来说，博弈论以及立足于其上的微观经济学，在观念上是假设博弈者像是（as if）理性地行为的。理性预期学派的创始人穆斯在回应对理性预期假设的批评时讲，“鸟不懂空气动力学，但鸟照样飞得很好”，所以我们研究鸟的飞行可以假设鸟像懂得空气动力学一样。

鸟的飞行能力是进化的产物。同样，以均衡作为博弈的解概念，可以把博弈的均衡策略看作是长期学习和进化所形成的稳定状态，至少是局部稳定状态。而制度由于它是与技术、资源一起决定博弈的结局函数，而制度的被实施和绩效是通过制度要求并预见的社会目标成为均衡策略的结果。所以，我们可以把可观察的制度看作是一个社会进化过程中的局部稳定态。

从进化的角度看，许多形形色色的制度，作为可观察的一种历史或现实存在，就像是冥冥之中有一个设计者的精巧建构。因此，机制设计理论的很多结论，从（局部）稳定的角度看，可以解释很多制度的出现。

机制设计理论的成功使人们考虑能否把它运用于制度变迁问

^① (Geanakoplos, 1994); (Dekel & Gul, 1997)

题，即，在规范意义上，如何进行一个演进社会的制度设计。但是，要把机制设计理论作为指导实践的规范理论，却并不像表面看来那么简单。因为如前述，从研究者的角度看，机制设计理论是存在性的，它只表明研究者作为一个系统外的客观观察者，可以确信存在一个机制以实现某些目标。但怎么把系统外的判断转化为系统内的^①？或者说这样的机制在现实中是如何被设计出来并得到实施？任何试图处理此自我相关问题的努力，最终都取决于理论在多大程度上是对现实的“恰当描述”。

当我们说存在的制度就像是精巧设计^②出来的时候，我们是一种事后的解释。而试图设计制度是一种事前的努力。显然，只要我们承认我们是生活在一个非严格决定论的世界里，我们就不能从根本上避免事前与事后的一致性。另外，我们也已提到，所谓的设计者是研究者，实施者是所有的博弈者，那么理论中设计和实施者的角色就不能简单地理解为现实社会中通常被称为制度的设计和实施的组织或个人。

现有的机制设计理论成果表明，只要我们选择合适的博弈解概念，几乎任意的社会选择规则都是可执行的。但问题恰恰在于：当我们讨论机制设计的时候，是假设逻辑上有无穷多种机制可供选择，一个社会技术上也能够随心所欲地选择制度。但是，我们从历史中发现，一个社会可选择的制度却是非常有限的。

机制或制度，作为一个博弈形式，其主要内容是策略空间和结局函数。而机制能够得到实施，依赖于这样几个前提：社会选

博弈论基础的很大一部分工作就在于此。

本文几乎没有涉及这个所谓设计的“合目的性”，即社会选择规则是如何确立的。一个简单的思路是，借用纳什规划（Nash program）的观念，我们可以把社会选择规则看作是社会谈判过程的结果，而谈判又可以看作是非合作博弈的结果。

择规则作为制度目标是协调的，制度执行者关于结局函数的承诺是可信的，以及人们有足够的理性来完成这个博弈。在博弈过程中，策略是人们在信念下的行为选择，而均衡策略则构成一个信念和行为的自我完成（self-fulfil）系统。也就是说，一个人相信别人会按照均衡策略作出行为选择，从而他也就选择了自己的均衡行为，而别人也是这样考虑的，由此均衡得以达到。所以制度及其所要求的行为规范首先应该成为人们的一个“普遍信念”（common belief）。

更重要的是制度，不管是自我实施的，还是被实施的，如果能够在很长的时间内保持稳定性，必然是意味着，人们在制度提供的博弈形式中的均衡策略也是足够稳健的，人们能够在不同的环境中，不因一时一地的得失而偏离制度所要求的均衡行为模式^①，制度也不会因小概率的偏离行为而形同虚设^②。所以，从实证的角度看，我们在大的时间跨度内能够观察到的制度必然具有很强的稳健性，同义反复地说，稳定下来的制度就是长期学习和进化过程中具有结构稳定性的均衡状态。

那么，这种稳定性来源于什么呢？我们所知道的历史上以及将来会发生的制度变迁又如何理解呢？我们可以把制度看作机制设计者利用其知识或技术作出的决策，而决定机制设计者的那种

比如，重复的囚徒困境博弈。

② 在理想的完全执行中，人们是不会偏离均衡的。进入非均衡路径，即“违法”行为，可以看作是博弈者理性程度不足造成的，比如关于自己因违法会受到惩罚的预期概率很小，而后者又与执行者的知识和技术水平有关。好的制度未必追求完全执行社会目标，但至少应该能够近似执行。所以社会惩罚犯罪主要目的不在于惩罚本身，而是在于表明制度执行者的能力和作出承诺，以改变人们的“非理性”信念。

更基本的制度又是其设计者的知识的产物^①。所以，我们可以把制度看作是人类社会应对环境不确定性的一种知识。这种处理使我们发现两点：第一，由于制度选择与个人决策之间的相似性，可以把制度作为一种特殊的知识来处理，制度的演进与知识的发展遵循着几乎同样的动力学机制；第二，由于机制设计者必然是社会的一部分，所以要逻辑清晰地分析这个问题，我们必须考虑制度分层^②。一个制度的稳定性和合理性原则上是由社会的更基本的制度和实施制度的知识所决定的。

某项制度一旦被选择，也就成为人们的知识或信念的一部分。那些稳定下来的制度以及相关的知识，给生活在制度下的人们以相当强的信念，使他会预期到自己的行为选择的后果，不会轻易选择制度所不允许的行动方式，更难得主动去改变现行制度。这种信念又会由于人们对其已有经验和知识的依赖而得到强化，制度和制度调整下的行为就这样形成一个自增强（self-reinforcing）的“鸡生蛋、蛋生鸡”的循环机制。而这个动力学机制有两个重要特点，即所谓的路径依赖（path-dependent）和锁定（lock-in）效应^③。

路径依赖使得最初的随机选择影响了后来的整个过程。在存在多重均衡的情形下，哪一个均衡成为现实本质上是对称破缺的结果。而结构稳定的均衡的性质决定了，由于随机的初始选择而

机制的两个重要成分，策略空间和结算函数，本质上都是由博弈者和执行者组成的社会的知识和技术水平所决定的。每个人的可行策略集，由他所掌握的知识和技术所限定。结算函数反映了两部分因素的结合，一是由技术决定的各种策略组合导致的物理性的结果；二是由制度执行者所施加的对比结果的调整和影响。后者原则上也是由制度执行者的知识和技术水平决定的。

② 当然，我们知道知识也是分层次的。

③ Arthur, 1990。实际上，阿瑟最初关注的是技术进步过程中的路径依赖和锁定效应。

一旦进入一个稳定状态，那么就很难脱离它^①。

这就有两点值得强调，一是由于人们对自然、社会、人生的思维方式以及由知识和技术所决定的生活方式的区别，每种社会的具体制度选择也各有差异，并且这种差异会被逐渐放大，使得不同社会的结构形态和实际运作千姿百态。特别在文明发生之初，所得到和形成的知识模式的差异对此文明有着持久影响，其中一个重要因素是地理环境^②。二是一个社会的政治、法律和经济制度本身也是人们的社会知识甚至自然知识的体现。我们所观察到的西方社会，其宪政制度之卓有成效，一靠制度设计精巧细致；二靠操作者理性运作，依赖于知识进步与科学昌明的成分是很大的^③。并且，好的制度对科技进步反有推动之功，两者互为循环，互相促进。为什么民主宪政制度与近代科学革命之肇始为西方所独专，重要原因在于，此两者有一个重要的共同因素，那就是继承自古希腊的逻辑严密的演绎分析和近代以来形成的经验实证的科学精神。

对于一个演进社会而言，这种路径依赖和锁定效应，使其制度和制度知识成为一定意义上的局部知识。因为这个社会没有也不能历尽所有逻辑上可能的世界，其主要的知识都是来源于它的历史和它存在运行的“小世界”。这就大大增加了学习其他文明成果、推动社会制度变革的难度和变革过程中的波折。那些微小的变革努力在这种封闭社会中，只会被消弭于无形。社会结构和制度的突变和跃迁，只能取决于其内部的强烈震动或外部的巨大冲击。

① 因此我们会发现古今中外，在许多外部观察者看来属低效率的制度竟然有着持久的生命力。

② 随着知识的进步，地理环境决定的痕迹会越来越淡。

③ 用黄仁宇的话来说，西方社会是用数目字管理的社会。

（三）社会进化与制度演进的动力学。

我们强调制度的稳定性并不是否认发生制度变革的可能性。这里有必要讨论一下哈耶克“合作秩序的扩展”的概念。哈耶克把西方社会与市场经济相关联的一整套制度的发展过程理解为“人类合作的扩展秩序”^①。那么，这个扩展着的秩序是如何在传统的基础上扩展的呢？与此相关的是，粗看起来，非西方社会欲学习西方的自由市场和多元民主制度都表现为发展中的不连续，是一种原来某些秩序的中止，哈氏的逻辑似乎很难解释清楚这种学习的可行性。

我们的知识和制度都是长期进化的产物。我们能够生存下来，除了运气或偶然之外，还在于我们知道如何运用知识实现某个目标。并且，随着环境的变化，进化的压力要求我们能够判断并决定（这种判断据以作出的知识是一种元知识）哪些知识是该丢弃的，哪些是该坚持的，哪些是该学习的。也就是说，我们的知识区分为基本的与派生的，前者是相比较更稳定的，后者附丽于其上但有相对的独立性。在现实里最基本的知识是那些直觉性、本能性的东西^②，是历史赋予我们且经过时间考验而得以保持的稳定知识，它对我们的约束像自然规律一样^③。而最基本的制度就是由最基本的知识所决定的“自我实施的制度”。它们赋予表层知识或制度以合理性，因为后者必须与它们相协调。改变那些派生的知识或制度是简单的，而改变相当深层的更基本的东

^① Hayek, 1988。

^② 我们在呼吸和思考以前是不必考虑是否该呼吸和思考的。

^③ 哈耶克认为，“自发社会秩序立基于下述两种规则：第一种是先天的，遗传继承的关于人的行为的普遍规则，它们形成于人种的生物进化的过程；第二种是习得的，文化传承的关于人的行为的规则。”第一种构成所有社会科学不言而喻的前提（Hayek, 1973）。

西的困难则要大得多，那需要更大的环境变化、更多的时间和学习另一种新思想理论的引入。或者说一种新观念的被接受是一个长期的学习过程，足够长的时间会使它们最终能够跳出传统和习惯的束缚，或者说原来的均衡状态被打破了。

简单地分类，知识进步和制度变迁有两种形式。

第一，在原来的知识体系中，某个经验性命题被修正了，它所造成的博弈者的决策选择也随之发生变化。在知识分层中，这种知识是附着于某个基本知识之上的表层知识。它与更基本的知识相协调，但并不是更基本的知识蕴涵着它^①。我们所观察到的经验性现象，从统计意义上给我们以信念，使我们相信这种知识下的决策是应对某类局面的最优策略。而当我们观察的经验性现象的有关数据足够多和足够广时，可能会改变关于环境的信念。我们有可能选择另一种知识，只要那种知识与基本知识依然是协调的。相对的，如果一个制度的环境发生了变化，例如，由于技术的进步，博弈者的策略空间和结局函数也会发生变化，或者制度执行者的知识和技术的变化使其可利用的信息和手段也有了变化，原有制度对博弈结局的调整结果可能已经不在社会目标集合中了。这时，它就有可能被新设计的制度所取代^②。

第二，原来知识体系中没有的命题被发现了，博弈者的策略

如果它被蕴涵的，那么应该与蕴涵它的知识是一个层次的。当然，由于我们不接受“逻辑全知”假设，这种蕴涵关系可以是一种受复杂性限制的蕴涵（人们相信基本前提，也就相信那些他们能从基本前提推演出来的命题）。在这个角度上，表层知识相对基本知识的协调性中包含着博弈者信念的成分。

^② 这是从制度设计者的知识改变的角度来说的。由于制度涉及到至少两人的社会，大部分非帕雷托式的变革都会遭到制度的既得利益者的反对。变革能否实施就取决于制度执行者的能力了，比如他可以用强迫（反对者如果不接受就会得到更坏的结局）或诱导的方式。在有些时候，反对变革者是弱势群体，强制性的制度变迁似乎容易做到，如二战以后日本的宪政改革。

空间也扩大了，所谓创新就是如此。所以，真正的创新本质上应该体现不连续性。创新如果是由过去所完全严格决定的话，那不是创新，我宁愿称之为知识扩散和制度传播。真正的进步必然包含跳跃或跃迁性。创新往往是产生自局部的突变，逐渐生长并推广到全局，而这个扩散过程既有决定论的一面，又会夹杂进随机性。随机性源于，对创新者而言，其创新行为是主观博弈均衡策略行为，这些行为中的哪些在最后成功是事前难以预测的，它们在竞争中会出现分化。决定论源于我们在学习和进化中的简单规则，即马太效应。人类社会的大部分制度也是这样，在基本制度之外，博弈者以试探和学习的方式达成某个协议或承诺，随着他们从遵守行为中获得好处，协议和承诺逐渐形成惯例和规范，甚至在得到基本制度的确认以后会成为法律等强制性的显性制度。

我们强调三个看法。

第一，进步意味着从可能性成为现实性，从逻辑上的存在性被推进到构造性。所以，哈耶克所批评的建构主义的“理性狂妄”其问题不在构造性本身，而在于缺少一种从元理论角度自省的清醒态度。构造性制度的边界来源于人类社会知识的局限性，即它受进化过程限制的局部性和受复杂性约束的有界性。

第二，基本知识和基本制度能够赋予表层知识和制度的变化以合理性基础。同时它们的相对稳定又保证了博弈者决策和社会生活的基本结构的稳定性。

第三，人类社会原则上由其知识和技术所决定，鼓励知识创新和技术进步的制度就尤其重要。制度关于个人决策权的安排只能是以两种方式进行：一是构造性形式，以直接或间接方式保证某些事前可描述状态的可行性（同时否定了相对的状态的可行性）；二是存在性形式，对制度设计者和执行者不能预见的机会不加限制，而给个人留出自由选择的空间。如哈耶克所言，“如

果有一些全知全能的人，如果他们不仅知道那些影响我们获得现在想要东西的因素，而且知道我们将来的欲望和需要，那么自由就没什么地盘了。为了给不可预见和不可确知的事物留有空间自由是不可缺的。所有自由的制度都是对这个基本的无知的事实的适应，以处置机会和可能性而不是确定性”^①。由于所有的创新都是来源于个人，所以越是实现了个人自由的开放社会就越有创新和发展的动力。这样一个制度的确立，是不可预定的制度演进过程中最值得设计的成分^②

二、关于第三批案研究的评论

（一）非正式制度与制度演化。

按照最初的设计，本案例集想围绕着非正式制度和正式制度的关系问题来描述和讨论中国正在进行的制度变迁过程。实施的结果是，有几个案例正面涉及到这一问题，特别是陈剑波的文章，提供了一个很好的案例分析。借此，我们想进一步讨论一下这一问题。

在制度经济学的文献中，不仅要一般地讨论非正式制度及其与正式制度的关系，而且要考察它们在制度变迁过程中的相互作用。这是两个既相互联系又相互区别的问题。如果说前者的讨论相对较多，那么，后者的研究则就不够充分，演进经济学在这方面有了一些前进，而现实的制度变迁过程可以提供很多进一步的思考。案例研究的意义正在于此。

^① Hayek, 1960, p29.

^② 我们可以尝试根据机制设计理论的一般原理给出几个演进社会机制设计的原则问题：多目标之间的折中；个人理性与激励相容约束；试错与学习导致的过程复杂性；机制必须具备的相对简单性与稳健性。