

钟学富

著



论思维过程的 物理机制

看物理学如何“入侵”心理学

中国社会科学出版社

钟学富 著

论思维过程的 物理机制

看 物 理 学 如 何 “ 入 侵 ” 心 理 学



中国社会科学出版社

图书在版编目(CIP)数据

论思维过程的物理机制：看物理学如何“入侵”
心理学 / 钟学富著. —北京：中国社会科学出版社，
2017. 5

ISBN 978-7-5203-0660-7

I. ①论… II. ①钟… III. ①思维科学—关系—
物理学—研究 IV. ①B80-05

中国版本图书馆 CIP 数据核字 (2017) 第 143855 号

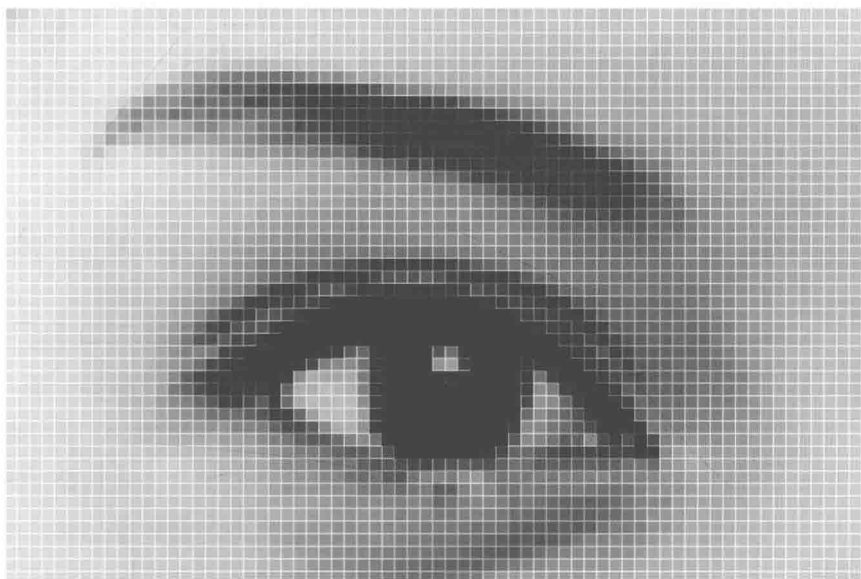
出 版 人 赵剑英
责任编辑 慈明亮
责任校对 闫 萃
责任印制 张雪娇

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010-84083685
门 市 部 010-84029450
经 销 新华书店及其他书店

印 刷 北京君升印刷有限公司
装 订 廊坊市广阳区广增装订厂
版 次 2017 年 5 月第 1 版
印 次 2017 年 5 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 23.25
插 页 3
字 数 368 千字
定 价 98.00 元

凡购买中国社会科学出版社图书，如有质量问题请与本社营销中心联系调换
电话：010-84083683
版权所有 侵权必究



图形意涵源自基础变异度的关联：在眼睛及类似的平面和立体图形中，基础变异度是每个像素的三原色（红黄蓝）的比例（决定视见的颜色）和总的亮度（光强）。设定每个原色有 10 种强度比例，则三原色共产生 1000 个变异度或 1000 种视见颜色，又设总亮度有 10 个刻度，则每个像素的总的变异度为 10000。假如每个像素的原色比例或视见颜色和亮度都呈混乱分布，平面上将不会出现任何图形。必须各像素的基础变异度严格限制或互相关联（每个像素的取值将不再独立，而依赖其他像素的取值），方可呈现出某种图形，不同的图形关联也不相同。例如眼珠内的像素须大体均呈黑色，眼白部分须大体呈白色，否则将不成其为眼睛。注意眼睛的图形还只是在几何及物理层面，不到心理的层面，须将眼睛的图形同记忆中的相关图像进行辨识或认知方面产生眼睛的意涵，过渡到心理的层面。辨识或认知是图形的外部联系，也就是新特性（心理意涵）只表现于它物（人脑）的外部关系中。正如电视机的“用途”只表现于人的关系中一样，没有人，哪怕电视机的各种零件连接正常，功能正常，能接收播放的图像和声音，仍然谈不上任何“用途”。

前 言

物理学从来不是也永远不可能是孤立的。作为最基础的实体科学，除了发展自身，它还有义务解释一切物质现象，因此它必须深入其他现象领域，哪怕最复杂和最高级的社会运动也不例外。这并非物理学家好管闲事，而是基础科学的职责所在，因为任何物质系统归根结底都是物理系统。疏通各个现象领域，构造严格、统一的世界图景乃是人类认识的最高境界，凡是科学家都该义不容辞。不过，实际上物理学迄今只给化学（原子的化合和分解）奠定了比较坚实的理论基础（量子化学），大体说明了生命现象（划清了生死的界限，解释了活机体为什么具有一定的自我治愈能力），还没有认真触及人的思维和社会运动。十多年前，笔者尝试将自组织理论用于社会系统的构建，以演绎方式解释社会的层次结构如何经“自由度缩并”产生，由中国社会科学出版社出版了两本专著：《物理社会学——社会现象演绎理论的探索》（2002年）和《社会系统——社会生活准则的演绎生成》（2007年），系统表达了笔者在这个方向上的努力。物理社会学不同于孔德的“社会物理学”，后者大体只是在社会现象中寻找某些物理关系的形式类比，由于类比推理太过弱势，可靠性很低，不能拿它当真。只有老老实实把自组织的理论概念用于社会系统进行演绎，方有可能真正获得合乎科学规范的可靠结论。这点意思笔者曾专门著文《什么是物理社会学，它有什么用处？》给予解释（见《科学对社会的影响》杂志，中国科学院主办，2009年第2期）。由于社会生活是通过人的自觉活动实现的，必须以思维运动为前提，在上面两本专著中笔者都指出，尽管现在还不能透彻了解思维的物理本性，但确定可以排除其中不包含任何神秘的要素，即便最困难的意志自由也可以纳入一个“选择模型”的物理框架，因此人类社会的产生完全是自然的物质过程，我们能

够合理地把对社会结构的演绎进行到底。可以认为,非平衡过程的热力学为结构和秩序的产生提供了一个普遍概型,既可用于社会系统,亦可用于人的思维。社会生活不是杂乱无章的,思维运动当然也不是,它们的结构和秩序都应当或必须纳入这个概型,即由外部信息源输入的信息流在系统内部结构中产生的信息积淀。

思维现象的研究极其广泛,从感觉、记忆、思考到情感、动机直至人格精神,而意志自由是全部问题的核心,比一般模拟性的人工智能或视觉、听觉的流行研究要复杂得多。人作为“能思考的主体”最主要的特性是它为什么能自主地作出一个又一个的思想和行为抉择;所谓学习、创新能力又是怎么回事?这当然最终要靠将来的神经生理学和认知心理学向分子水平的发展,但在此之前物理学必须给出一个粗略却不失完整的概念框架,用技术的语言来表述思考的主要过程和机制,否则各种具体研究便将失去规范,可能堕入歧途。常常见到各种发明具有“独立意志”的机器人的报道,它不光是非专业的媒体语言,也夹杂专业人士的期许。这就触及非常原则的问题:人工智能有没有边界;界限应当在哪里?回答这些问题不能靠猜、靠推测、靠个人睿智,而是要从普遍的科学原理去推论、演绎,所得结论才不是凭空杜撰,才有可靠的依据。演绎和推论可能出错,但这并不可怕,因为演绎过程是公开的,错误一经发现谁都可以纠正。

但理论思考不能只靠原则的推论,它需要实验的先导。脑电波的发现是神经生理学的一项重大成就,通过置于头部的电极可以记录人在不同思考和情绪状态下脑电波的图形,以外部检测对思想的内容作出某些客观的判断,这已成为当今探索人的神经活动的最重要的实验手段之一。因为只有正常状态的活人才能思考,对神经活动的实时检测必须是无损伤的(不能随便开颅),所以这项实验受到诸多限制,不能进到颅内置入电极,虽然它提供的信息有限,仍然有着难以替代的作用。脑电波来自神经元活动中的电化学过程,它包含大量的神经脉冲,尽管每个脉冲的能量水平很低,整体上仍足以造成宏观的检测效果。这项发现和神经解剖学和生理学的其他成就一道,确凿证明了心理活动的物理、化学基础,将其还原为自然的物质过程,所以具有重要的判定作用。这项成果很可能具有重要的实用前景,假如能从脑电波的图形确切分辨思想的内容,那就可以借助它实现所谓“意念控制”,比如飞机和车辆的驾驶,只要事先在计算机中存储

与向上、下、前、后、左、右等方位的意念对应的电波图形，通过对比识别这些意念，进而控制飞机或车辆的运动。在紧急情形下，如发生交通事故或空战中驾驶员受伤，丧失肢体控制能力后仍能由意念操纵飞机或车辆安全降落、行驶到适当位置。这个过程在物理上完全合理，与江湖术士玩弄“特异功能”所谓的“意念控制”毫不相干。困难只在分辨意念不是轻而易举的事情，但它毕竟可以看作是朝着“读出思想”的方向迈出了最初的、可喜的步伐。

神经解剖学证明人脑的组成没有任何怪异，全部属于门捷列夫周期表上所列的化学元素，有着正常的分子、原子构造。没有怪异的物质就不会有怪异的过程，更别提任何神秘的因素了，脑神经的全部活动本质上都属于已经被确凿了解的物理、化学运动，不过有自己特殊的表现形式罢了。无论思想现象多么复杂，最终都只能靠普遍的自然规律来解释，正因为如此，DNA 双螺旋结构的发现者之一克里克才斗胆提出了“灵魂的科学探索”的口号，而分子生物学家更是“毫无忌憚”将自己的实验手段尽可能广泛地直接用到神经活动的各个领域。

但是，对科学而言光凭信念和勇气是远远不够的，我们需要一个符合规范、与已经确认的科学原理相互匹配或“接轨”的、普遍的概念框架来指导意识的产生和思维本性的研究。例如在生命起源的问题中，我们不光要弄清每个物种如何在特殊的 DNA 分子控制下生长和发育，而是首先要划清生与死的界限，说明活物和死物的本质区别。此外，通常活机体都具有一定的“自我治愈”能力：桌椅坏了不会自动修复，而皮肤划道小伤口不久便能愈合，它是生命的特征，所以理论必须给活机体的“自我治愈”能力以合理的解释。现在这些都已在不可逆过程热力学的基本原理中找到了满意的解答。对思维运动来说也是一样，我们首先需要一个基本的概念框架。机器可以动作，却不会思想，人不但会动作，而且有思想，这个“有思想”是什么意思？它就牵涉到如何定义“智能”。可以指出，通常基于“图灵测试”给出的定义^①并不合于严格的科学规范，它抹

^① 图灵测试给出的智能定义只是行为主义的，有许多不确切和人为的因素，本书正文第一章的讨论将指出，学习和创新活动在智能的概念中具有重要的区别，必须依据这些能力来理解智能的概念。

杀了创新活动在智能发展中的核心作用，必须彻底改换思路（第一章）。进一步，人都有“自我”的感觉（主体对自身存在和活动的认知），能行使个体意志，对每一件事情作出自己的思想和行为抉择，它们是基于什么样的物理过程，须用怎样的概念给予准确的描述，这些才是问题的核心。研究思维本性必须首先抓住这些根本，不能仅仅局限于各种心理现象，如感觉、记忆、判断、推理等的神经生理学机制，尽管这些问题在分子生物学的水平上都各有其重要性。

意识和生命现象显然联系十分紧密：只有活机体才能思维，而思想发展始终和生命发展相生相随，思想的器官是生命运动的指挥中心，高度智慧的生物同时也是最发达的生命体。从认识的角度看，两个问题在尚未弄清时都曾使人感觉神秘，有过许多悬念。而今生命现象已经大体得到说明，所以回顾一下历史，看看人们当年是如何克服那些似是而非的概念壁垒，如学习“过来人”的经验一样，对于考察意识产生或思维本性可能有重要的启发。面对生命这样的复杂现象，人们都会去猜测它背后的原因，但猜有各种猜法。一种是直接采取人为的、貌似合理的假定，认为它来自某种“活力”。这其实就是神秘主义，因为谁也不知道什么是“活力”，更无从检测，等于只取了个名字，对认识深化毫无助益，反而阻碍了探索的道路。

另一种则是努力把未知现象纳入既有科学理论的框架，用已知的概念和原理去解释它们，这就是科学主义。这样做常常会遭遇困难，因为在每个具体历史时期认识水平总有局限性，既有的理论框架可能不足以说明新的现象。但越在这个时候越不能动摇，需要静观和等待新的概念和理论出现，以迎接将来的发展。对于生命现象，人们曾经困惑于它和热力学定律的表面矛盾，因为克劳修斯的第二定律断言孤立系统的发展是趋向平衡，而生命的特征恰恰在于相对稳定的结构，无论动植物，还是微生物都有外部的形态学结构和内在的生理结构，并且这些结构在生命的整个存续期中不断生长和发育，显然不平衡。这个矛盾直到 20 世纪中叶不可逆过程热力学出现之后才算弄清，原来生命体属于“开放系统”，和外界有着物质和能量交换，因此不断接受外界熵流或信息流（“吃进”负熵），造成所谓“非平衡定态”。在这种状态下，由于不平衡，所以容许系统拥有各自的结构（包含能量耗散的“耗散结构”）。又因为是定态，所以结构会有

相当的稳定性，当发生小偏离，出现伤病时，如内分泌轻度失调或皮肤、肌肉的小损伤，系统的发展是自动恢复原状：失调会回归正常，伤口会逐渐愈合，甚至无须药物帮助，这便合理解释了机体的“自我治愈”能力。当然这种能力相当有限，如果偏离定态太远，使稳定性条件破坏，系统便将走上另外的发展道路，不再回归先前的定态：严重的伤病会夺走机体的生命，使其死亡。借助非平衡定态的概念解释了生命的基本特征、划清了生和死的界限，就将它还原为基本的物理、化学过程，或者把生命看作特殊的物理、化学现象，如血液循环属于黏滞流体的力学现象，消化、营养吸收则属于特殊的化学反应一样。这些概念今天已毫不新颖，它们早成了当代医学的基础，到医院看病时，进行各种化验、检测是临床诊断的一部分。在这个意义上，生命体已经获得了百分之百的自然本质，就像“机器”一样执行着新陈代谢、生长发育、生息繁衍的基本功能，不再有半点神秘主义的影子。不同的只是，通常的机器是金属、塑料等材料搭建的静态结构，而生命体则是蛋白质、脂肪、纤维素等搭建的耗散结构或活结构，但两者都具有承载过程结构的能力，即允许可执行的程序结构（软件）在其上运行，充当硬件的角色，因而地位上完全可以相互类比。

从热力学观点划清了生和死的界限，同时也就看出两者之间并无不可逾越的鸿沟。人们通常习惯于把机体的活结构和没有自我治愈能力的静态结构（如桌椅）截然分开：桌椅一旦损伤不会自行复原。然而，静态结构或死结构之所以不能自我治愈只是由于切断了同外部信息源的联系，一旦恢复这种联系，死结构照样可以被“激活”。桌椅损坏不能自动复原，人却可以修复它们，只要把人和相关的工具、材料纳入一个大系统，这个修复仍然有权叫作“自我治愈”。修复过程受头脑控制，所以系统在修复时接收了外界的信息流，完全符合开放系统的基本原理。只是注意在热力学上要将人的头脑和身体分别划归不同的系统，头脑作为信息源单独归于“外界”，这叫“脑体分离原则”。社会生活的秩序：道德、法规，都是由人建立的，也是人工秩序，一旦遭到破坏也可被人的意志纠正，同样属于自我治愈。必须参照同外部信息源的关系来看待死结构和活结构，这才与热力学的概念框架保持一致。

现在面对意识和思维现象，情形完全类似。怎么去猜测它的原因？一种办法就是假定存在不死的“灵魂”或“精灵”主宰人的一切、控制人

的行为。人死了，失去自我的感觉，是“灵魂出窍”，人们在感觉困惑时突然冒出“灵感”，使人豁然开朗，则是上帝在传达旨意或神灵在发出启示等。这无异于走“活力论”的老路，结果不言而喻，所以决不可取。然而遵从科学主义则须审视当代科学是否真的创造了必要的前提，足以给这类比生命更复杂的现象（因它在生命的基础上产生）设计出合理的概念框架或模型，提供系统的科学解释？可以指出，在非平衡过程热力学和相关领域的发展之后，这个问题的回答是肯定的，它就是本书要做的事情。

我们需要一个合适的平台，用它来构建“能思想”的系统，亦即智慧生物或人的范本。现在最复杂的个体系统是“自主系统”，它不仅会按照预设程序做规定的事情，而且有一定灵活性，可以根据实际情况自行调整做法以达到更好的效果。这要求它部分模拟人的特性，能够存储信息（记忆）和决策（挑选后续的发展道路）。高级的自主系统甚至还有“学习”能力，即自建某些程序指令，可以说是最接近人的一类系统。不过和人相比，它们仍然存在很大的差距，除了静态结构和耗散结构的不同之外，核心的区别是所有程序包括学习都是事先设计好的，必须人要它做什么它才做什么，要它学什么它才学什么，不会主动“自学”，所以远不及人独立和灵活，更不容许创新，也没有自我的感觉。无论多么先进、复杂的自主系统都不能认为具有独立思考的功能，因而并不真正“自主”或具有独立意志。就像电脑虽然运算速度等方面强人许多，整体上却远逊于人类。然而，仔细分析这类系统却能发现，即使在现有科学概念的框架下，它仍然有重要的扩充和发展余地，自然界之所以创造了人类这样的智慧生物，正是充分挖掘了这些潜力。我们探索思维的本性就是要弄清自然界从哪些方面做到了这点，揭开了这个秘密，就能懂得“意志自由”和“自我认知”究竟是怎么回事，就能明白为什么不能在人的意志指挥下去刻意建造具有“独立意志”和“创新能力”的机器的道理。当然，本书叙述的扩充或改造只是理论的演绎，并非真的去搭建实际的系统，因此可以避开一切技术细节和困难，设想和发现这些细节是将来科学发展的任务。不过我们要求演绎的每个步骤必须逻辑通畅，合乎科学规范，不能和现有的理论冲突，以任何标准衡量都是完全合理的。

自主系统最重要的是它的反馈控制中心，如计算机的操作系统，其结

构决定了它的功能。所谓“结构”包含硬件和软件，前者指中心的物质构成，如机器多半是由金属、塑料等无机材料的部件搭建的，而生物个体则是基于脑神经的解剖学结构，它的主要组成材料是蛋白质。至于软件则是指信息的存储（留下记忆痕迹的步骤），还有与各种功能对应的特殊程序和规范，以及演算加工的信息，两者互相配合以保证加工时能准确、便捷地调用。简言之，中心必须保存工作目标和实施步骤的详细信息才能控制工作部分的运作，就像上级政府必须保存完整的法律、法规条文才能监管下级机构和公民的行为，考察他们是否尽责、有否违规一样。巨大的信息量使得这个子系统高度有序，并且需要保持高度稳定。尽管如此，却必须看到，迄今没有任何自主系统的反馈控制中心，除了人的外部监管，还能自己监督自己，靠自身的力量改善和提高其功能。这很容易理解，因为中心是人设计建造的，属于人工结构，一旦建成便被精心呵护，不能损坏，也无需改进。即使改进也是“下一次”建造时“人的任务”。换言之，机器的自主系统其反馈控制中心在建造之时就排除了“自主进化”的可能，任何系统一旦建成便切断了与人脑这个外接信息源的联系，变成了失去发展活力的死结构。

问题的症结正在这里，人们可以期盼一个永恒的静态结构，自然界却不允许。即便物质的最小单元，夸克或基本粒子，现在知道其性质也依赖所处的环境，不会绝对“全同”或始终如一，何况具有复杂内在结构的反馈控制中心，岂可是僵死和固定的？既然如此，就必须使中心的功能和结构都有演化的可能，而且演化必须是真正“自主的”，即由中心自行选择外部的信息，利用它们对系统的运作和发展实行控制，以保证系统整体在特定环境条件下的顺利发展。为什么必须实行“自主”？因为没有任何外部信息源能保证对系统的发展长期（更不消说永远）实行“正确”的控制。实际上，系统周围的环境始终在变化，充满了随机性，系统的发展要求其行为能恰当因应各种变化，趋利避害，这个责任不可能依靠外部信息源或外部控制担当，因为任何外部物体本身就不可能长期不变，并和系统维持固定的关系，只能由反馈控制中心通过“自我提升”去承担。

那么中心如何提升自我，或者说如何生长和发育呢？就软件而言，主要是增加记忆或存储的内容，不断累积知识和经验，补充应对外部事件的程序结构，产生新的行为模式。这需要学习，并且要创新，即自主地编制

新的程序结构，扩大应对外部事件的范围。然而现行机器的学习方式是不足取的，因为它是被动、受人控制的，叫它学什么才学什么，叫它怎样学就怎样学，而主动学习是沿着系统运行的轨迹进行，自主选择学什么和怎样学，更可以创新，即不仅模仿已有的成熟方案，还同时创建前所未有的新方案。就人而言，生活轨迹从不固定统一，包含许多随机性，这是因为环境复杂，谁也不能严格预期何时何地会遇到什么事情，因此无法一切照旧规则行事。这要求人的行为具有自主性，实行独立决策，在众多可能的方案中进行选择，或者说行使个体的“意志”：意志的本质就是选择，是通过实行选择产生或形成的，没有选择便无所谓意志。生活轨迹不能严格预期，因此人能获得什么本领也不能严格预期：谁能对科学前景和未来社会作出准确的细节判断？按照定义，学习是模仿构建“已知”的程序步骤，而创新则是尝试“未知”或前所未有的程序步骤。但两者都同样在控制中心构建工作程序，在这个意义上创新也是学习，只不过不遵从既定的模式，需要针对随机出现的情况去探索或尝试有效的应对方式。随机性对于创新活动是非常实质性的：谁也不能预期何时、何地、何人能创出何种“新”来——谁曾预期“相对论”恰恰是德国一个犹太裔年轻人，时任专利局小职员的爱因斯坦在1905年创造出来的？其实“未知”一词逻辑上便蕴含着随机：既然未知，便有多种可能。不过学习和创新未必都真正适合系统的利益目标，未必代表先进和进步，完全可能是落后和倒退：人既可做好事，也可做蠢事和坏事，犯罪手段也在不断创新。这导致历史发展中的曲折，虽然它整体上仍在前行。

既然如此，便产生了对中心决策实行“监控”的必要：必须合乎系统发展目标的决策才是正确的，而错误的决策将可能危及系统的安全和利益。实行监控的只能是中心自己，而不是更高的监控中心——否则永远存在最高的监控中心由谁监控的问题。这便导致一种“自反馈”，即反馈中心将自己的物理状态报告给自身，使中心在运作的每时每刻都“知道”自己在做什么，这便是“自我”意识的起源，对自身决策行为的历史记忆便构成每个人的自我心理环境。需要指出，自反馈是自然界的一种基本行为模式，也是导致各种非线性效应（物理之上的各种高级运动形式都有各自的非线性效应，化学中有自催化反应，学习中已有的知识激发求知欲，银行的复利，社会上贫富差距的扩大等）的主要机制。这个概念从

信息和物理的角度来看是完全合理的，因为思考中的“信息加工”归根结底是系统内部物理状态的变化，而各种物理状态正是信息的唯一载体，当然可以被传输，并反馈到自身。自反馈和从系统对外工作部分而来的反馈信号并行不悖、并不矛盾，两种反馈信号，依赖于承载的信息内容，均可可是各种各样的，诸如此类的细节需要留待将来分子生物学水平上的神经解剖学和生理学去探讨证实。给反馈控制中心加上自反馈，使其实行自我监控，系统便有了“决策”或行使意志的前提，否则决策不受限制，等于随机决策，早晚一定会使系统自毁。有了决策能力，中心便可以不断调整系统的目标结构和实施步骤，进行创新，这是一个实质的进步，也是人或智慧生物区别于机器的主要标志。

硬件是软件运作的平台，软件的扩充似乎可以置于任何硬件的基础上，包括静态结构在内，没有原则的证明无论什么样的静态结构都不适合于构建包含自反馈的控制中心。但问题是，任何静态结构的系统或子系统都不能保持永恒，真的僵死固定，因为不可避免有内部老化、外界侵蚀、运动磨损等机制，使塑料变脆变软、金属生锈断裂，导致系统连接松散，最后坍塌，这是一个熵增的自发过程，难以防范和绝对避免。自主系统的反馈控制中心也不例外，自然界一定会有使其老化、受损的机制，使其寿命有限。唯一可以长期、甚至无限制延续的只能是活结构或耗散结构，因为这种结构具有一定“自我治愈”的能力，足以应对小的扰动，通过生长和发育可以抵御或弥补自发的熵增，维持结构稳定不变。不仅如此，即便扰动很大，或多次扰动的后果累积使系统最终坍塌，但开放的、与外界存在物质交换的耗散结构还可拥有“自我复制”的能力，或者说活的系统能够生息繁衍，实现世代传承，因此，尽管个体寿命有限，但作为物种的基本结构信息依然保留在后代身上，就群体而言并不消失。看看人类已在地球上生活了若干万年，迄今仍绵延不绝，身体的基本结构始终保持，显然没有任何静态结构的机器可能达到如此漫长的寿命。哪怕设想地球毁灭，也不能排除人类集体搬迁到另外的星球生活的可能。在这个意义上，人体的生物信息，虽然会有变异或演化，其基本部分理论上甚至能无限期地延续！

既然软件要生长和发育，不断扩充，承载它的硬件也无法交由静态结构担任，复杂的软件只有足够复杂的硬件才能承载，两者需要同步增长，

只有活结构才能通过生长发育以至进化给不断扩充的软件提供足够的活动平台。生物进化证明，动物的脑容量，尤其皮层面积随着进化不断增加。大象的脑子重过人脑，皮层皱褶却很少，远逊于人类，因此信息容量不足，没有足够的载体作为信息加工的舞台，这说明它的智力何以会低于人类。当然脑的发育不是孤立的，而是与身体其他部分有着不可分割的联系，这些道理是很容易理解的。

现在可以说基本弄清“人”或智慧生物的本质了，它不过是一个“超级”自主系统，特点是反馈控制中心不能僵死和固定，可以生长和发育，具有学习和创新的能力。这里所谓“创新”是指在包含随机性的条件下，通过思想的自组织构建新的行为模式和程序结构，为此系统需要反馈控制中心通过自反馈实行自我监控，不过无论对内或对外的思想监控，归根结底都是物理的控制，只能通过物理信号如神经脉冲的传递来实现。具有以上特性和功能的自主系统就是智慧生物的范本，人在其中不过是发育较为完善罢了。理论上这似乎并不复杂，但它经得起推敲、质疑、考验吗？首先是常识的检验！要害当然是反馈控制中心的生长发育，但它其实很自然。试想，神经系统和活机体联系在一起，在机体生长发育的同时它岂能自我隔绝、纹丝不变？换句话说，活机体绝对容不下一个“死脑筋”！现在知道头脑的生长发育分为两个部分，一部分由遗传基因或 DNA 规划，另一部分却不是。前者决定脑的基本结构，人在出生后不久就形成了，且在一生中基本保持不变。神经元很少更替，但形状却不断变化，它们互相联系所依靠的枝枝丫丫——“突触”和“树突”在外部信息影响下随时都在生长，造成新的神经网络，淀积着新的外部信息，为反馈控制中心的生长和发育提供了可靠的物理支持。回忆耗散结构的两种基本类型：与外界只有能量交换和同时兼有物质交换的，前者可以“起死回生”，后者却不能（因交换物质后系统组成变了）。因人体吃进食物属于后一种，所以寿命有限，但人却可以自我复制或生育繁殖，使物种的结构信息得以保存。神经系统也一样，神经元作为活细胞需要正常的营养吸收和物质代谢，包括氧的供给，能量则通过线粒体提供。个体的头脑是注定要死亡的：现代医学把“脑死”作为比心跳停止更主要的死亡判据，然而亲子依然拥有完全的头脑，继续执行反馈控制中心的功能。

这也同时回答了人工智能的一道重要难题，即人能否设计和建造拥有

“创新”能力或独立意志的机器？答案无疑是否定的，否定的理由并非技术能力的限制，这类限制总可逐步克服，今天的能力不行，明天技术发展了则有可能。相反，这种否定是原则性的，因它包含一个悖论：建造机器本是为了实现人的控制，而建造一种“谁都不知道它能干什么”的机器恰恰使人失去了对它的控制。这种系统只能以耗散结构的自我复制，或生物学方法获得，那就是人的后代。为什么单单只有这样的自主系统能被容许和接受呢？这是因为亲子作为后代，身体和头脑的结构由于遗传是大体类似的，允许以“普适性的教育手段”去规范他们，使其创新活动沿着正确发展的方向进行。相反，设想一群来源各异、五花八门的创新型机器或智慧生物，便很难采取统一的教育手段，由此产生什么后果完全不堪设想。这也同时牵涉到人工生命的问题，因为机体和神经的生长发育需要同时在遗传基因中规划，迄今人工生命都是从现成的基因培养活的机体，并不包括人造基因，显然这个目标大可质疑，因为设计生命必须纳入神经系统，等于设计创新型机器。

各个要点的细节将在正文中阐述，但这样一个以“超级自主系统”为轮廓的智慧生物模型使我们确信，当代科学已经提供了理解意识和思维现象的必要概念框架，不仅不需要神秘主义，也没有必要胡乱猜测，钻进早期存在主义那样的“我之为我”的死穴，把人搞得糊里糊涂，只消老老实实以科学的方法深入研究，把自主系统放到复杂的外部环境中，使其各部分（尤其反馈控制中心）自然进入生长和发育的程序，因而需要实行决策、需要自反馈的监督、并且需要活结构来承载便足够了。正如划清了生和死的界限，“活力”等根本不存在的虚拟概念就自动被扫进历史的垃圾堆一样，从宗教教义衍生而来的“灵魂”之类也再无立锥之地。

探讨意识的产生或思维的本性，重点是考察主控中心意志的行使。对人而言就是所谓的意志自由，即作为独立的思考单元，人为什么可以以及如何对每个“当前事件”作出自己的思想或行为抉择，它显然最有实际意义。确认了意识或思维现象的自然本质后，这个问题就必须回归物理层面，以神经元和神经网络的几何和物理结构为基础，通过其上神经脉冲的激发、传递等活动为基元过程进行解释。但意志控制的是高层次活动，有语义解读和形象表达，不是静态的神经网络或动态的神经脉冲传递本身。各种语义或形象乃至肢体动作都表达某种“思想意涵”，代表着主控中心

的意志或意愿，相对于几何和物理特性，它们是全新的信息。这表明必定有系统的层次过渡发生，即由几何及物理的层面过渡到心理的层面。基元的物理过程属于“前意识”阶段，它们内容复杂，因其牵涉众多神经元的集体行为，且每个神经元的物理状态都不单纯，包含许多变化的可能性，或者说“自由度”很多，不像思想意涵那样相对单纯。在由脑电波主导的运动意念控制中，“向上”或“向下”是很简单的指令，可对应的脑电波形状却很复杂，而这种虽然复杂却不失规律性的波形需要大量神经元的协同动作，或由一组特定的神经脉冲的传递来实现，它对牵涉到的每个神经元的物理状态和行为都有严格的要求，否则将影响整个脑电波的波形，导致畸变，使“向上”或“向下”的意涵变得模糊。其他更复杂的思想意涵也有类似的情形，这表示从几何及物理层面过渡到心理层面，产生思想意涵的新信息时必须包含“自由度的缩并”，甚至不止缩并一次，它是系统层次过渡的普遍规律（附录五）。缩并的原因，一般说来都是由于基础自由度之间产生了“关联”，造成“自组织运动”，我们将见到，神经活动中的确存在这样的自组织过程，原因很简单：严格、准确而又灵活、细腻的各种思想意涵不可能由杂乱无章的神经结构或神经过程来提供。我们将据此解释思想意涵或语义的由来，虽然现在还缺乏相关神经生理学的细节，无法分析层次过渡的具体步骤，但这并非原则的困难。思想意涵是基本的心理要素，思考都是在意涵的基础上进行，控制也是意涵的控制，不是控制个别神经元或神经脉冲。各种思想意涵造就个体的“心理环境”或“内心世界”，全部心理学都是在分析和研究它们，只不过我们现在要将这个研究推进到微观。

由于缺少相关神经生理学的细节，我们无法实现从基础变异度到心理意涵的具体编码，那是未来进一步研究的任务。为了强调这个困难属于“技术”性质，可以对比一下普通电脑中存储器的物理状态与其存储的信息内容之间的对应关系。当今二进制存储器的基本单元有0、1两种状态，它们可以组合为十进制的数码，然后再组合为字母或其他复杂的符号、词语、图形、故事。组合的实质是使基础变异度之间实现特定的关联，因而减少了实际变化的可能性（相当于自由度缩并），并获得特定的信息内容，这个过程通常叫信息的“编码”，它规定信息内容与基础变异度之间的对应关系。由神经元和神经网络，以及神经脉冲的组合到语义、图像的

层次过渡与此完全类似，同样是基于神经结构和神经过程的组合或关联，只是复杂性大大增加了。每个神经元的位置、形状、与其他神经元之间的联结状况都可能对编码有某种影响，而且神经元本身就有复杂的结构，它还参与包括外部信号的传输以及营养吸收的生化过程，过程本身又有各种特性，从中分析和确认哪些几何及物理要素作为基础变异度是极困难的，找出其间的关联就更复杂。但思维运动毕竟是有秩序的，思想意涵作为控制信息更不能含糊，普遍的因果律要求意涵是能分辨的，基元的神经结构和神经过程也必须能够分辨，否则我们就不能指出语义或图像存在差别的物理原因，导致无原因的意识现象。这个原因只能在基本层面上（神经元、突触和树突、神经网络的联结、神经元内部的生化过程、神经脉冲等）寻找，于是我们借助“心理—物理平行主义”的概念：两个层面，一个心理，一个物理，它们并行不悖、互相对应和依存，所以叫心理—物理平行主义，实质是确认心理意涵具有物理编码的可能。之所以采用“主义”一词是因为对于智慧生物，例如人类，编码的具体过程或方式尚不清楚，远不如计算机的0、1那么简单。从事记忆研究的心理学家对这个问题的难度应当深有体会，但不论技术难度如何，编码的可能性是毋庸置疑的，所以用“主义”来表达对这种可能性的信念。脑电波的语义解读正是这种对应关系的证明，电波图形的差别当然是由于神经元和网络运动的物理差别造成的。

既然意志的抉择说到底还是物理运动的结果、物理规律的选择，那么意识或思维活动就该纳入从力学运动到化学反应、再到生命现象同样的自然过程的序列。这些运动全都遵从动力学的因果律或“广义的变分原理”^①：实际的运动轨迹总是在容许的“变分轨道”（路径、方法）中使相关“泛函”取极值的那一条。因果律和变分原理本是同一件事的不同表述，因为有极值的唯一性，所以才有“一定的原因产生一定的结果”。本书用“选择模型”来解释意志自由正是基于这一思想，以便把因果律引入一个

① 数学上函数依赖于点的坐标，而泛函则将“点”扩充为一条轨道，所以它依赖于轨道的形状。由于轨道是通过函数来描写的，故泛函也称为“函数的函数”。寻求泛函极值的专门数学分支叫变分法，而物理学的“变分原理”则断言，真实的物理运动轨迹是使泛函取极值的那一条，所有的基本运动方程，如牛顿方程都可以从相应的变分原理推导出来，它们和变分原理物理上是等价的，同样表达了因果律的要求。