

科学与人文

## 关于意识的哲学思考<sup>①</sup>

金观涛

**摘要** 意识研究中存在着两大基本范式,这就是行为主义范式和哲学范式。哲学范式主张唯有意识才能理解意识;行为主义范式把一切解释建立在受控实验之上,力图在意识的解释中排除灵魂。本文通过二十世纪至今的研究回顾指出,这两种对立的范式中任何一种都不能理解意识。意识的科学解释有待于这两大范式之综合。本文提出意识研究的递归方案,认为意识的科学解释和我们现在熟悉的“科学解释”不同,它会导致唯物论和唯识论界线的消失。

**关键词** 意识的科学解释 心脑模型 梦和神经网络 递归解释

笛卡尔认为,我能怀疑一切,但不能怀疑“我正在怀疑”,否则,一切确定性都不存在。据此,“我”可以等同于“我思”。但是,“我”真的能用“我思(有理性)”来界定吗?早在两千多年,佛学已证明“我”不是“我有”。用“我有”来定义“我”,就是把我视为一种存在的属性。然而,只要自我意识尚未消失,即使再蒙眬,我都知道自我的恒常性;今日之我即昨日之我。而“我有”是流变的,“我有”显然不是“我”。但取消了一切“有”,独立于“我有”之“我”又是甚么呢?

### 一 意识的科学解释能排除“意识”吗?

哲学家把意识比喻成隐藏在云端中的高山,不仅人类至今尚未认识它,即使力图去仰望山顶,亦会产生巨大的眩晕感。有人称意识为认知领域的最后一个问题,它或许是科学研究的极限。然而,今日盛行之对意识的解释恰恰是在刚过去的二十

**作者简介:**金观涛,香港中文大学中国文化研究所讲座教授,现任台湾政治大学讲座教授。

<sup>①</sup> 本文曾在2009年2月1日中国国际神经科学研究所(China INI)举办的“医学与哲学研讨会”上报告。在写作过程中,吴立新先生曾为本文提供资料,特此致谢。

世纪建立的,它包含哲学和科学两种完全不同的范式。

1900年弗洛伊德(Sigmund Freud)《梦的解析》出版,奠定了以精神分析为代表的哲学范式。该书提出两个基本观点:第一,梦的本质是愿望的实现;第二,做梦是潜意识存在的证据。“精神分析”意味着对意识的解释发生了自笛卡尔以来革命性的飞跃。从此,非理性、潜意识以及自我的不同层面和形态都成为意识内省和思辨的对像,“自我”的研究开始和精神病治疗以及心理学相联系。弗洛伊德的理论冲击了心理学、哲学和社会科学,甚至对文学艺术都产生了巨大影响。我们可以称其为意识研究的哲学方向。

与哲学解释并行不悖的,是大脑和思维探索的行为主义方向。如果说,意识解释的哲学范式是从人人自明的意识出发,通过思辨和内省,以认识“无意识”、“潜意识”和自我的各种形态;而行为主义范式则完全相反,其目的是为了在意识解释中赶走哲学思辨的鬼魂。行为主义把原来只有通过内省定义的意识之特性统统用行为模式来取代。例如一只狗被汽车撞了,下次见了汽车就会害怕,通常人们是用狗有记忆来说明这一点的(心理学称之为生物对外界刺激的敏感化),行为主义认为,狗有记忆和其行为模式改变等价。因为行为模式是可以通过观察来研究的,甚至可以定量地加以测量,故行为主义开启了心理学科学化、量化的转向。

1950年代以后,行为主义范式开始和控制论、系统论互相交融,结下了丰硕的思想果实。其中最重要的就是用行为的结构(黑箱输入与输出的关系)代替本质,使得有机体的行为(如对内稳态的调节、适应、试错和学习)均成为可以用机器模拟的东西。神经网络、人工智能研究兴起。从某种意义上讲,今日人类在计算机仿真和人工智能中取得惊人的进步,都缘于这一范式转化带来的革命。行为主义反映在意识解释上,就是图灵(A. Turing)对思维的定义。图灵指出,如果我们对机器提一系列问题,当其回答和人没有区别时,即可认为机器和人一样能思维。<sup>①</sup>

令人深思的是,到了1980年代意识研究的上述两个方向似乎都走到了尽头。虽然,今天研究大脑已离不开计算机比喻,神经网络和程序是其分析基础,越来越多的科幻电影甚至在描绘当计算机具有意识时人类命运会如何,但科学家终于发现,意识对计算机至今仍是不可思议的。确实,再复杂的人工智能和专家系统都不可能通过图灵试验,没有一部超级计算机具有自我意识的迹象。<sup>②</sup>换言之,一个世纪的研究

① 表面上看,图灵提出的是思维行为主义定义,而没有涉及意识。显而易见,我们亦可以用图灵试验给出意识的行为主义定义。

② 至今为止,没有一部计算机可以通过图灵实验。某些表面上接近图灵实验的结果实际上都不是严格意义上的测试。如1993年11月,美国波士顿计算机博物馆曾聘请了10位没有受过计算机训练的市民通

究得到一个结论:仅仅依靠行为主义研究纲是不能理解意识的。

仍以记忆为例子。心理学发现,以行为模式改变来定义的记忆实际上只是记忆的相当低级形态,它只能刻划反射性或程序性(而不是反思性)的记忆。人学会游泳但事先并不知道在水里应如何动作,这确实可用行为模式改变来代表;但我们一般说的记忆,则是指主体通过回忆,可换起过去的经验。心理学称之为陈述性记忆(declarative memory)。它对生物之所以重要,乃是唯有依靠它生物才能主动地去适应环境。更重要的是,只有具有大脑(而非一般神经系统)的生物才具有陈述性记忆,它具有符号性、建构性、会被遗忘等特点[史奎尔、肯戴尔 2001, 页 143—209]。主体追溯这类记忆很像搜索引擎登录那样进行检索或对过去进行重构。显然,理解陈述性记忆必须假设主体的存在;也就是说,意识的科学解释似乎不能排除意识。

这样一来,意识作为大脑属性的探讨似乎仍要回到大脑的哲学思考上来。但是,正如行为主义已穷图末路一样,以往盛行的精神分析亦不能进行下去了。二十世纪哲学最大进步是科学哲学的确立。学术界公认,任何一种涉及真实世界的(科学)理论必须具有可证伪性,否则理论是不能深入发展的。而弗洛伊德学说似乎不可证伪,甚至有人认为它是伪科学。

今日意识研究的进步深入除了其理论可被实验证实或证伪外,还要求这些实验观察超越行为主义,即它必须加上基于研究者对意识体验之上的分析。也就是说,需要哲学方法和科学方法实现某种综合。二十世纪上半叶制定的意识研究纲领之所以不能继续下去,是因为哲学方向和科学方向的互相分离。事实确是如此,近三十年来对意识研究最重要的进展恰恰都来自于这两个原来分离方向的交汇点上。

## 二 梦的实验研究

这方面最重要的例子是关于梦的实验研究。梦的实验研究和弗洛伊德内省式的分析不同,它首先必须确定人在甚么时候正处于梦境中,以便进行受控的观察和测量。以往这是不可能的,但 1953 年,阿瑟林斯(Eugene Aserinsky)及克莱特曼(Nathaniel Kleitman)一项关键性发现使其成为可能。这就是发觉人在睡眠中有时会同醒着时一样作快速眼球运动(rapid eye movement,简称 REM),他们将这种睡眠

---

过 14 分钟问答交谈,判别参赛者是计算机还是人,提的问题是人际关系和大学生活等,居然有 4 个软件骗过了至少 1 位裁判。这是否意味着计算机已接近通过图灵测试?并非如此,关键在于,计算机的回答是基于人对问题的答案制成,机器只是根据提问中的关键词对号入座。科学家一致认为,要完全按照图灵的设想不规定话题,人工智能不可能通过测试。

称为快速眼球运动睡眠(REM睡眠)。1957年,狄曼(William Dement)与克莱特曼把处于快速眼球运动睡眠的人唤醒了191次,发现了157次是有清楚的梦境记忆。其它睡眠(没有快速眼球运动,NREM)中的人被唤醒160次,只有11次有做梦的报告[拉维2002,页101]。该实验迅速被其它人重复。人在睡眠中看到梦境一定伴随快速眼球运动,终于作为一个基本事实被确定下来了。<sup>①</sup>

一旦发现快速眼球运动和做梦之间的必然联系,睡眠和梦的实验研究就大踏步进展。首先被重视的是“何为看见”的研究。人们发现先天的失明者的梦没有形象,亦没有快速眼球运动。而一个人失明越久,他睡眠时快速眼球运动就越少。换言之,因任何看见(包括在梦中看见)都必定有快速眼球运动,科学家开始研究梦境和快速眼球运动模式的关系。<sup>②</sup>由于可以在试验者发生快速眼球运动时叫醒他,并清楚地把梦境记下来,使得做梦在某种程度上是受控的。这样,就可以对梦进行更详尽更系统的记录,并设计“梦和记忆”、“梦和学习”以及“做梦是否不可缺少”等种种实验。而且,动物也存在快速眼球运动睡眠,<sup>③</sup>研究者可以用动物实验观察做梦时大脑的化学变化。1960—1990这三十年,梦的实验研究取得了极大的进展。

其中两个结论也许对揭示梦的本质具重大意义。第一,梦在统计上有明确的模式,它随年龄、性别不同而不同。而且,每个成年人的梦在类型上有相当的确定性[罗柯2007]。但就每一个具体清醒的梦境而言,其展开完全是随机的,不存在任何叙事逻辑和意义。人们可以对梦境进行任意切割后再进行随机重组,研究者根本不可能发现哪些梦是经过重组的。故有人据此认为,这表明佛罗伊德对梦本质的断言很可能是错的。梦不一定是愿望的实现,也很难说是潜意识存在的证据[霍布森1999,页142—144]。然而,因弗洛伊德学说中“愿望实现”和“潜意识”的含混性,证伪和证实都很难,心理学家对此陷于不休的争论中。

第二,实验发现做梦时大脑神经介质浓度的宏观分布明显和清醒与无梦睡眠(NREM)不同,它有点类似于人丧失理智发疯时的状态。第二点格外重要,它导致脑化学状态和心灵状态关系的一系列新研究,是精神分析不能想象的。为了说明这

① 这里所说的做梦,是指明确地看见,下面称为有清晰的梦境,而非指人的意念和意识活动。有关梦的定义本来是含混的,但自从REM睡眠发现后,梦和梦中看见(梦境)明确区别开来。除了特殊的脑受伤病人,当发生REM睡眠时,都有明确的梦境,这一点我们在第五节有关看见时详加讨论。

② 如斯坦福大学睡眠实验室所记录的“乒乓球之梦”。观察到一实验者一连串26次从左到右、从右到左的眼睛移动,将其唤醒,要他报告梦境。他说正在梦中看乒乓比赛。但在一般情况下,REM睡眠中眼动很复杂,至今没有发现其模式和梦境的对应关系。

③ 此处动物是指哺乳动物。爬行动物没有REM睡眠。对于哺乳动物,如果没有快速眼球运动,是看不见任何东西的。快速眼球运动在进化链何处出现,是一个极有趣的问题。一些动物(如青蛙)只能看到运动之物,快速眼球运动用眼自身运动使得静物可以看见,这是一大进步,它是做梦的前提。

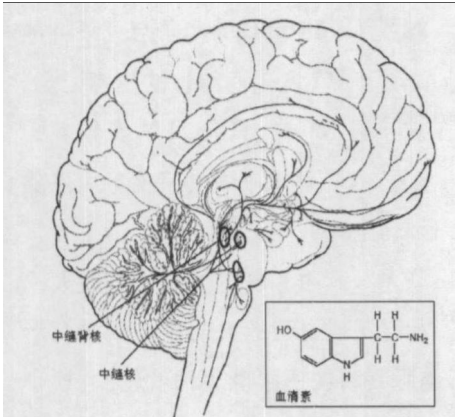


图 1

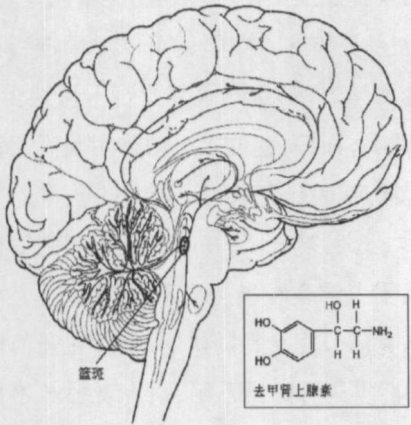


图 2

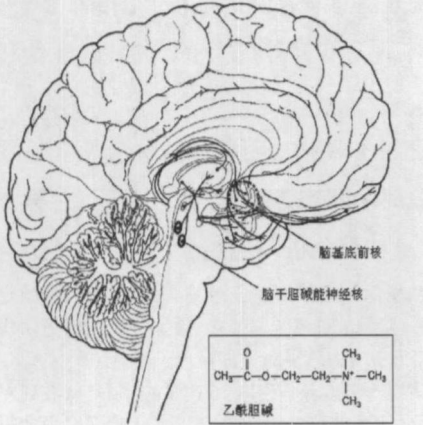


图 3

一点,必须解释一下何为神经介质。

我们知道大脑和神经系统的任何两个神经元都不直接相联,一个神经元的电脉冲(兴奋或压抑)必须通过神经元之间的化学物质传导给另一个神经元(使其兴奋或压抑)。这些化学物质就是神经介质。神经介质分两大类,一类是胺。它包括血清素、多巴胺、去甲肾上腺素等。另一类是碱类,如乙酰胆碱。因为一个神经元的电脉冲如何通过神经介质将信号传给另一个神经元过程很复杂,且各类神经介质起作用的方式随神经网络不同亦各不相同,故以往人们只重视神经介质的微观状态,从未考虑到它们在大脑中的浓度宏观分布和人的精神状态的关系。

那么,大脑整体的神经介质宏观状态是由甚么决定的呢?不同的介质在大脑不同部分合成,通过神经元簇(脑干核或腺体)分泌到大脑各部分,影响到整个神经系统(使大脑有关部分浸在其中)。就拿胺的主要成份血清素来说,合成血清素(方框内是其分子式)的神经元位于脑干中线神经元簇(称为中缝核)。它们投射到脊髓、脑干和小脑,包括丘脑,丘脑下部,一直到边缘系统和大脑皮层(图1)。<sup>①</sup>

合成另一种神经介质去甲肾上腺素(方框内是其分子式)的神经元位于包括蓝斑核的几个脑干核,它们的轴突延

① 本文从图 1 至图 12 及其相关叙述都取自:Hobson 2001.

伸到脊髓、脑干和小脑,上达丘脑,丘脑下部,边缘系统和大脑皮层(图2)。

合成多巴胺(方框内是其分子式)的神经元则位于中脑,它们投射到边缘系统(中脑—边缘系统),大脑皮层(中脑—皮质通路),以及锥体束外运动系统(黑质—纹状体通路)(图3)。

对于碱类,亦是一样。合成乙酰胆碱(方框内是其分子式)的神经元位于脑桥、脑干和基底前脑。它从脑干核局投射到丘脑、丘脑下部、基底前脑和边缘系统。脑基底前核则投射到大脑皮层和边缘系统(图4)。

梦的受控实验发现,清醒时大脑主要处于两种胺(血清素和去甲肾上腺素)控制之下。一旦睡着,胺的浓度就开始减少。当人做梦时,乙酰胆碱浓度(还包括某些部位多巴胺浓度)不断增加;当意识深陷梦境不能自拔时,大脑处于碱控制之下。

显然,这些合成神经介质的神经元簇(脑干核或线体)是受大脑控制的,即何时在何种状态下分泌神经介质以改变神经介质宏观浓度,取决于大脑整体上处于清醒还是做梦等状态。这样一来,如果找到代表大脑处于清醒还是做梦等宏观状态的另一一些参数(变量),就可以用这些参数(变量)之间的互相作用刻划大脑精神状态。

众所周知,人在清醒、睡眠还是做梦时脑电波是不同的(图5)。因此,脑电波构成描述大脑意识的第二个重要参数。<sup>①</sup>此外,人在清醒时意识到外部世界存在,意识主要关注(或处理)外部信息。而在睡眠或梦中则主要处理内部信息。这样,处理

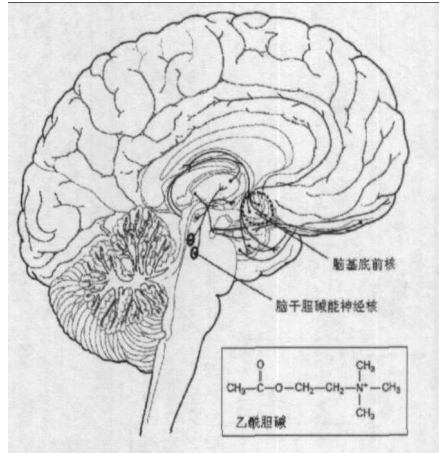


图4

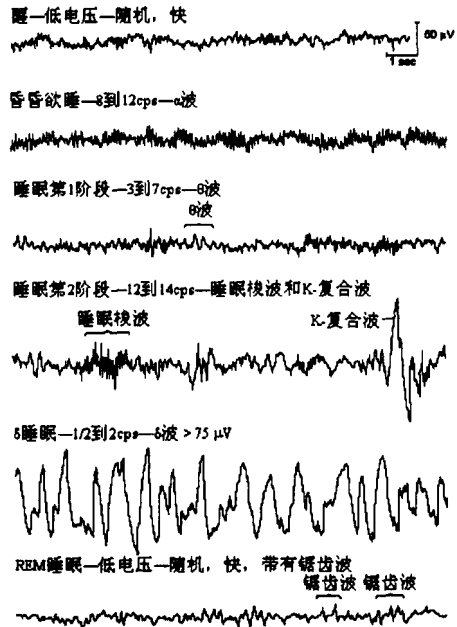


图5

① 脑电波 1920 年代末由德国贝尔格(H.Berger)发现,1935年就应用到睡眠研究上,定义从醒到睡的变化。

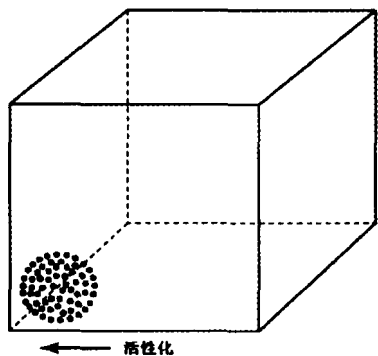


图 6

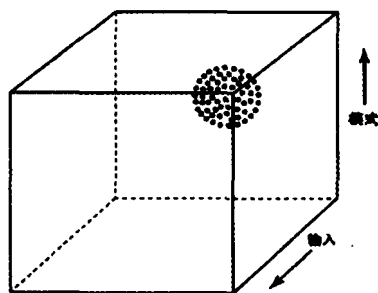


图 7

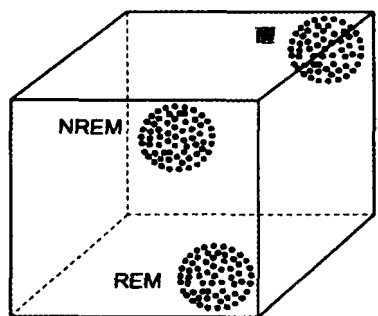


图 8

信息的来源构成第三个基本参数。利用这三个参数,就能用模型讨论意识的宏观状态。这方面最重要的工作由梦和睡眠专家霍布森 (J. Allan Hobson) 完成,他将其称为意识的 AIM 模型。

### 三 意识的心脑模型 (AIM)

早在 1950 年代艾什比 (W. Ross Ashby) 他的控制论名著《大脑设计》中,就设想用行为变量 (参数) 的组合 (相空间中的状态) 来代表复杂的生命系统 [艾什比 1991]。当时他没有办法把大脑的意识亦作这种表示。今天, AIM 模型利用三个参数做到了这一点。这三个参数第一个是 A,它是“脑电波强度”亦称活性化;第二个为 I,它是“意识关注的信息”亦称输入输出选择通道,第三个为 M,它是“神经介质宏观浓度”,亦称化学模式。

我们来看组成 AIM 的状态空间的三个参数的准确含义。活性化 (A) 为脑电能量坐标轴,它表征信息处理的能力,其大小取决于脑电波 (EEG) 平均振幅的倒数及网状丘脑皮层系统神经元的激发率。在图 6 中状态空间

的前侧面,从左到右意味着活性化 A 由低到高。第二个参数为图六中相空间的深度,它是一个代表信息来源的坐标轴(I)。其值大小意味着输入输出选择通道关闭 (I 较小,即意识关注系统内部产生的信息) 还是打开 (I 变大,即意识关注系统与外界交换的信息)。第三个参数是神经介质浓度 (M),它代表控制大脑的化学模式。在相空间中它用高度坐标轴代表,刻划胺类 (去甲肾上腺素和含于血液中的复合胺) 和类胆碱 (乙酰胆碱) 化学物质影响的组合。M 大意味着胺高 (上面,代表胺主导),M 小意味着胺低 (下面,碱主导)。<sup>①</sup>霍布森认为,心脑的各种状态都

① 严格说来 M 是由多个变量组成的矢量,有多少种神经介质矢量就有多少个分量。把所有介质用一个

可以用相空间的点来代表。<sup>①</sup>

首先,我们来看如何表示昏迷。昏迷时人没有意识,即此时活性化 A 水平低,胺作用 M 弱,外部输入选择通道关闭 (I 值极低)。故昏迷区域在相 (状态) 空间的前方左下角。当然,有多种昏迷的症状,比如睁眼昏迷和闭锁综合症,它们会占据状态空间左下角的不同区域 (图 6)。

当大脑处于相空间前方右上角 (图 7) 时代表甚么呢?这时 A 很高,胺主导大脑意味着人是清醒的,但 I 不大标志大脑正在处理内部信息,这代表人看到幻觉。同样,相空间前方右下角处 A 很高,但 I 不大以及 M 很低,该状态意味着大脑活跃但受碱控制,而外部信息通道关闭,即相当于人在 REM 睡眠状态或处于梦境中 (图 8)。

当大脑位于相空间后方右上角时,A、M、I 三个值都很高。I 高意味着大脑处理外部信息,加上 A 和 M 值高,显然这代表清醒状态 (图 8)。根据实验,大脑处于无梦睡眠 (NREM 睡眠) 时,A、I、M 的值都大约为清醒时一半左右,即它位于相空间的中心位置 (图 8)。

这样一来,所谓睡眠实为从“醒”状态到“非梦睡眠”状态 (NREM) 再到“做梦睡眠”状态 (REM),再回到“醒”状态的循环。换言之,正常睡眠在 AIM 状态空间中形成一个椭圆轨线 (图 9)。该曲线可以解释睡眠的阶段。众所周知,一个睡眠周期约为 90 分钟,一晚大约有 5 个睡眠周期。每个睡眠周期分若干类型和阶段,如浅度睡眠 (REM)、较深睡眠,更深睡眠,最深睡眠 (NREM) 等。这四阶段循环反复,每一个周期从 NREM 睡眠开始,再从深睡眠到 REM 睡眠,然后醒来。从一个睡眠周期进入另一个睡眠周期,转变的速率是不一样的。醒到 NREM 睡眠较慢,NREM 睡眠到 REM 睡眠很快,REM 睡眠到醒非常 (最) 快。因此,一夜做 4—5 场梦。

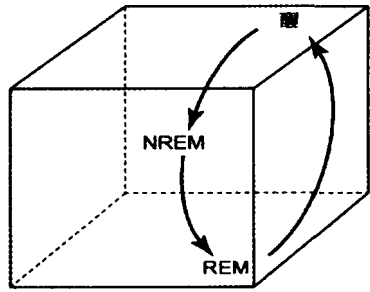


图 9

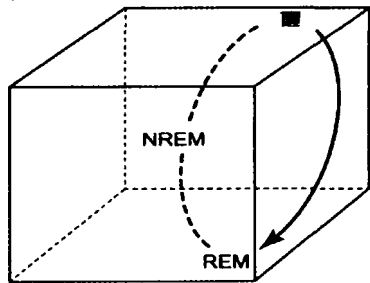


图 10

变量极不严格。但为了在三维空间将心脑状态表示出来,把矢量简化为标量。由此可见,AIM 模型把多维相空间简化了,人们发现霍布森模型很多缺陷均缘于此。

<sup>①</sup> 见 Hobson 2001. 图 7 至图 12 以及相应文字解释亦来自该书。



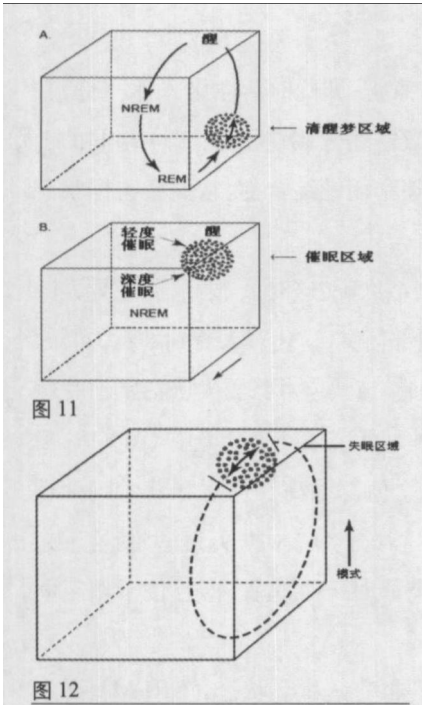


图 11

图 12

当睡眠不按该轨道运行,往往意味着发生非正常现象。例和在嗜眠发作时,入睡过程和正常睡眠轨线相反。无论是在白天还是黑夜当睡意来临之时,病人直接从醒进入REM睡眠。这是由弱胺和强类胆碱的混合作用所致,它导致昏睡状态的出现(图10)。

读者会问,相空间后方右下角(图11-A)又代表甚么状态?当该状态不稳定(或亚稳定)时,意味着大脑在做清醒梦。即被试者既保持部分REM睡眠状态,又恢复部分清醒意识。它为REM睡眠和醒之间的不稳定平衡。清醒梦很容易地回到REM睡眠或者醒来。如果这些状态是稳定的,无疑意味着人丧失理性,即以做梦的心灵处理外部信息,也就是精神病。

当相空间状态处于后方右上部,但I比清醒时小(图11-B),大脑处理的开始从外部转化为内部,这只有在催眠条件下才能发生。通常,催眠状态是不稳定的。它会转化为NREM或经REM醒来。当系统稳定地停留在清醒区附近时,意味着焦虑和失眠。焦虑拒绝入睡是因为胺作用很活跃,被试者徘徊于睡和醒的交界处(图12)。

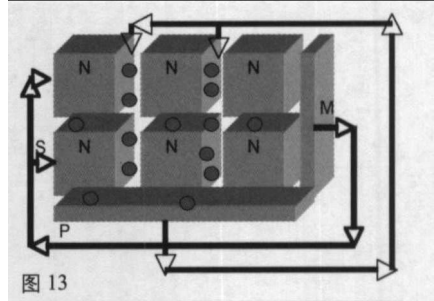
AIM模型能相当准确地把握脑心宏观状态,甚至可以涵盖很多精神病和睡眠研究的临床经验。在某种意义上讲,它是第一个关于意识的宏观模型。但是大多数研究者对它评价不高,因为它是唯象的。人们认为它只是经验之总结而不知所以然,即不能回答为甚么意识状态可以表示在相空间中,而且为甚么它们大多是稳态?更不能通过它理解意识的本质。

#### 四 新的综合:相空间的意义

我认为,AIM模型的意义在于它开始了新的综合,即终于把哲学思辨范式和(行为主义式的)科学范式结合起来了。只要我们分析组成相空间的三个参数,就可以发现其中两个(A和M)可以转化为大脑神经网络宏观模型中最基本的状态变量,而另一个参数I的界定必须依靠人对意识的内省理解。也就是前两个参数是

属于行为主义范式,而第三个参数是不能独立于意识,属哲学范式。该模型说明解释意识需要同时运用两种范式。为了讲清这一点,我们先讨论前两个参数。

行为主义范式最大的特点是用神经网络来建立大脑模型(将神经元的行为组织起来以



形成具有大脑某一功能的网络)。1980年代控制论学者 Heinz von Foerster 考虑到神经元的独特性,即两个神经元不直接相联,信号传递必须依靠介质(这和计算机有极大不同),建立了用这种独特部件组成的神经网络模型。他认为,大脑(神经系统)具有如图十三所示的整体结构。[Foerster 1984]

图 13 中一个个方块(N)为神经元,圆点代表神经介质。神经元通过介质互相联系,组成巨大的网络。神经网络控制运动肌 M 和垂体、腺体 P(当然亦包括合成神经介质的神经元丛、脑干核)。Von Foerster 认为该网络最重的整体结构就是封闭性,也就是网络的任何一种输出都反馈到系统本身,成为输入的主要成份。换言之,如果 S 代表神经系统电信号输入,M 为运动肌输出,那么,M 必定反馈回来成为输入,即 S 大致等于 M。这是第一个封闭回路。同样,介质浓度决定神经网络整体运行模式,它控制垂体、腺体以及合成神经介质的神经元丛和脑干核(P),但神经介质宏观浓度却是由 P 决定。这里,同样存在 P 反馈回来第二个封闭回路。

Von Foerster 进一步指出,对于这种整体上封闭的网络,存在的宏观状态均为本征态。本征态概念来自量子力学,量子力学中任何一个可观察状态都满足本征方式;因该状态在算符作用下不变,故称为本征态。二阶控制论将该结论用来表示神经网络的封闭性对其输出和内部状态的影响,即和任何一种封闭网络相关状态都是本征态(整个网络对状态影响相当于算符)[金观涛 2005, 页 31-54]。本征态分稳定态和亚稳态两类。这样一来,图 13 所示模型的神经网络电输出和介质浓度都应该是本征态。

在 1980 年代,二阶控制论从来没有想到用本征态来表示人的意识。今日将 AIM 模型和 Von Foerster 神经网络相比较,立即看到 A 即图十三的电信号输出,M 即神经网络另一个重要宏观参数介质浓度。至于它们必须是稳态(部分是亚稳态)恰恰和本征态属性一致!换言之,Von Foerster 神经网络居然可以成为 AIM 模型的基础。

读者或许会问:既然 AIM 模型中所有意识状态都是本征态,为何神经网络有这

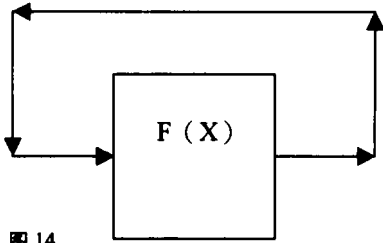


图 14

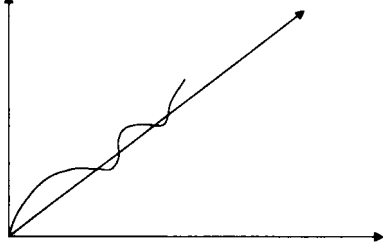


图 15

么多不同的本征态? 关键在于神经网络为非线性系统。系统论早已证明对于非线性(复杂)的封闭回路,本征方程存在多个解,即有许多本征态 [金观涛 2005, 页 217-223]。举一个简单例子可说明这一点。图 14 为一最简单的自耦合系统,轴出 Y 反馈回来作为输入 X。其本征态满足本征方程:  $F(X) = X$ 。显而易见,本征态由曲线  $F(X)$  和对角线 ( $Y=X$ ) 的交点决定。当  $F(X)$  是非线性函数时,交点不止一个(图 15),即非线性自耦合系统有多个本征态。

我认为,根据 Von Foerster 神经网络,或许可以解释为什么脑必须做梦,即 REM 睡眠为何对大神经网络不可缺少?很可能它是唯有通过介质才能互相联系的巨大网络系统的一个特征。如前所说,大脑和计算机不同,任何两个神经元必须靠介质相联系,即神经元只能通过释放神经介质来传递电信号,使网络中其它相应神经元兴奋或压抑。这样,介质浓度宏观稳态作为一种自组织系统存在之内环境对系统结构稳定不可缺少,但神经网络工作过程亦必定对介质宏观的稳态有所影响。正是这两种因素决定了内环境稳态必须周期性更替,它直接表现为睡眠和做梦。

发现神经元之间如何通过介质互相联系的机制乃为近几十年神经科学最大的成就。原来,神经元通过介质互相作用分为(电信号)快突触传递和慢突触传递两类。电信号快突触传递机制如图十六所示。先是神经元(第一个 N)动作电位打开其突触前末稍的离子通道,环境中钙离子的进入导致该神经元所储存的神经介质量子式地释放(一次五千个分子)。这些被释放出的介质分子立即和相邻的神经元(第二个 N)突触后末稍结合,造成环境中钠离子进入该相邻神经元,其后果是激发出一动作电位(根据阈值大小),使得电信号沿第二个神经(突触)元传递到其前末

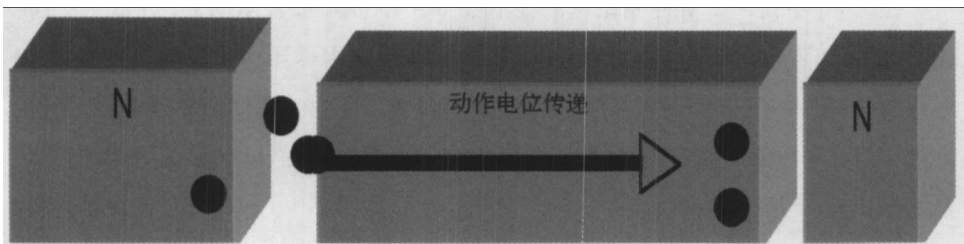


图 16

稍,如此等等。第二类(慢突触传递)不打开离子通道,而是传导介质引发一系列生化反应,使突触后末稍蛋白磷酸化或去磷酸化,以改变该神经元下一次对介质的敏感度(阈值)。<sup>[李葆明 2005, 页 59—88]</sup>

无论是打开离子通道还是使末稍蛋白磷酸化或去磷酸化,都是神经元对传导介质和介质存在的宏观环境的操作(包括调节和控制),它对该神经元及其所处的宏观介质环境必定存在微小的影响。当这些影响不破坏神经元末稍和介质宏观稳态时,整个神经网络能有效地工作。但系统论有一基本法则,就是这些暂无妨碍的副作用会积累起来,构成破坏整个网络结构稳定的无组织力量。我们称之为功能异化导致结构稳定性的破坏<sup>[金观涛 2005, 页 308—317]</sup>。这样一来,神经元末稍状态以及它所处的介质环境宏观稳态必须过一段时间后改变一次,以清除无组织力量。做梦和睡眠代表了和清醒不同的介质环境稳态,它们周期性地取代清醒意味着通过介质稳态更替以消除清醒时积累的无组织因素(它导致胺效能的降低)。这就是为甚么有大脑的生物均需要睡眠(做梦)而计算机不需要。<sup>①</sup>

请注意,仅仅用参数 A 和 M 组成的模型不能说明意识。图十三所示的结构只能证明存在着介质环境内稳态以及内稳态周期性更替的必要性,它至多是意识存在的物质基础,而不能说明意识是甚么。确实,分析该系统的总功能和子系统及其部件的功能,没有一个能说明意识。事实上,AIM 模型能解释意识所依靠的是第三个参数 I,以及它和另两个参数 A 和 M 结合的方式。

第三个参数 I 为注意力开关(对输入输出的选择通道),其大小的含义由大脑是否意识到外部世界的存在来决定。大脑时时刻刻在接受内部和外部信息,只有意

<sup>①</sup> 此处,我们只证明睡眠对复杂的生物不可缺少。至于做梦和睡眠的关系,目前还没有搞清楚,如注 8 所说,一些动物没有 REM 睡眠,故无法判定它们是否做梦。关于做梦对大脑必不可少的观点存在着多种理论。一种通常说法是“梦为本能的预演”,因刚出生的婴儿拥有大量 REM 睡眠,故认为梦和神经系统发育有关。而且,如果是当爬行动物进化到哺乳动物时产生了梦,做梦作为求生本能的预演说似乎更有说服力了。另一说为梦是人学习功能的重要组成部分。特别是短期记忆转化为长期记忆时,REM 睡眠起着不可取代的功能。因短期记忆是“情景式”的,胺控制是保持这些记忆的物质基础,而长期记忆是“结构式”的,属于碱控制下的行为模成,故短期记忆转化为长期记忆需要介质环境从胺控制转化为碱控制。该说因和 1980 年代克立克从“神经网络理论”推断梦由信息被大脑记住过程产生暗合,故西方大多睡眠研究者都认为无梦会破坏长期记忆的形成,特别是近年来它似乎可以解释为何日常生活入梦要滞后数天(6—8 天)。关于日常生活入梦滞后最著名的研究由朱费(Michel Jouvet)作出。他在波斯湾战争时进行测试,发现战争开始 5 个星期后,几乎一半的梦和战争有关。在一个人一夜做的 4—5 次梦中,当天的事常进入第一个梦,以后的梦往往和几天前的事有关。生活入梦滞后和学习有甚么关系?是最近常讨论的问题。关于这方面的例子可看最近出版的一本通俗读物:《我们为什么要浪费时间睡觉:梦的科学解析》。然而必须指出的是,因发现无梦睡眠人记忆不受影响的反例,关于梦对长期记忆不可缺少仍没有成为定论。但有一点无疑义,大脑如不睡眠(包括做梦)(即不从胺控制周期性地转向碱控制)是不能保持清醒状态的。

识到外部世界的客观存在,大脑才可能反复鉴别外部信息的可靠性并把它和内部信息比较以知道外部信息代表甚么。然而,何为大脑意识到外部世界的存在?这是必须凭内省思辨才能知道的,即I的意义界定必须依靠内省而不能用在外的行为可观察量。此外,AIM模型强调胺控制下大脑处理信息模式和碱控制下不同,但如何描述这种不同呢?换言之,人在醒着时处理信息方法和清醒的梦中处理信息方式的差别,也是只能通过内省才能知道的东西。

由此可见,AIM模型之所以能把握意识的不同状态,依靠的是内省因素和行为主义模型的结合。这是一种新的综合。

## 五 意识解释的递归方案

也许,读者对这种新的综合不甚满意,因为解释意识仍必须依靠对意识某种更原初状态的理解作为前提,它是半吊子不彻底(二元论式)的。更有甚者,其思维模式和常识唯物主义对意识的解释相当类似。唯物论把大脑存在作为意识的前提,然后用人人自明的意识来定义意识。前面我们用图十三的模型加上I等唯通过内省才能定义的因素来解释意识,这里,神经网络模型是常识头脑的细化,I和梦中处理信息模式和醒不同也只是常识性意识的细化,两者在方法论上似乎完全同构!

那么,新的综合究竟新在何处?我认为,只要我们认识到意识的解释不能排除意识,又要让该解释尽可能被实验证实,就有可能发展出一种将科学与哲学始终结合在一起的新方法。我称之为意识解释的递归方案。或者说,新的综合之所以为新,并不是指其一个个具体的结果,而是指它将对应一个可以展开的过程,该全过程才能对意识的本质作出解释。

简而言之,正是基于意识的解释不能排除意识,在意识的研究中,每一次都需要涉及两批不同性质的变量(参数)。一批是行为主义方法的变量(参数),我们记为B集合中的元素,它是可以用实验测量并用来建构科学模型的。另一批是只有通过人的思辨才能理解的变量(参数),我们记为C集合。在每一次解释中,必须用B集,因为只有这些参数形成的关系可以用实验证明,但单纯用B集合中的元素是不够的,我们必须附加C集合中的元素,附加之所以可能,这是因为我们有意识,可以用意识理解意识。事实上,正恰恰是基于可理解性,构成的整体模型才是合理的(虽然不是每一个细节都可以用实验检验)。

表面上看,因为解释中包含因素 C,对意识的科学解释仍没有最后达成。但我认为,只要采用一种类似定义递归函数的方法,可以将解释中 C 类因素不断化约,使其为 B 集合中的元素和更原初的意识因素之组合。即随着每一次解释向前推进一步,研究者必须注意采用 C 集合中不同的元素(避免循环论证),并去发现更为基本的元素。我们可以期待内省的意识因素越来越小。也就是说,只要意识解释的递归方案不断顺利展开。意识之谜或许有一天可以解开。意识解释的递归方案可以定义如下。

首先,对于必须解释的意识状态  $C(n)$ ,我们用比较原初的意识状态  $C(n-1)$  和行为主义变量  $B(n)$  建构一模型,使得  $C(n)$  得到解释。我们将该解释模型记为  $F$ ,于是有: $C(n) = F[C(n-1), B(n)]$ 。接着,我们必须进一步对  $C(n-1)$  作出解释,仍是采用  $C(n-2)$  和  $B(n-1)$  两类参数,得到另一模型  $K$ ,于是有: $C(n-1) = K[C(n-2), B(n-1)]$ 。这样一步一步推下去,得到一个由  $F, K$  等组成的解释链。如果在链的延伸中, $C$  集合中的元素和  $B$  集合的差别最后可忽略不计,二元论分裂的困难也就克服了。也就是说,或许意识的正确解释和我们迄今为止所熟悉的科学解释不同,它不是靠一个模型实现的,而必须依赖一个或几个由  $F, K$  等组成之收敛长链达成。

我们来举例子。如将 AIM 模型和常识唯物论作对比,无论作为意识载体大脑之模型,还是用于解释意识的意识原初状态的要素(如 I 的定义),都大大前进了。AIM 模型中涉及的意识内省因素为各种不相同的“看见”(“梦中看见”和“醒时看见”)以及“注意力开关”,如果常识对意识的内省定义为  $C(n)$ ,“看见”为  $C(n-1)$ , $C(n-1)$  比  $C(n)$  更具原初性。<sup>①</sup>那么,我们能否建立新的模型,把  $C(n-1)$  还原为更为原初性意识要素呢?

显然,醒时看见过程可用模型图 17

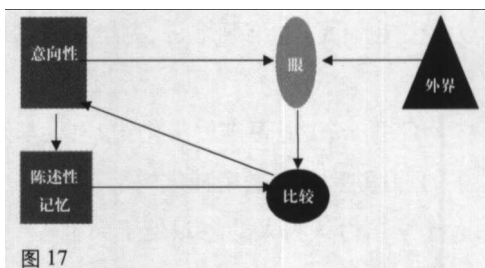


图 17

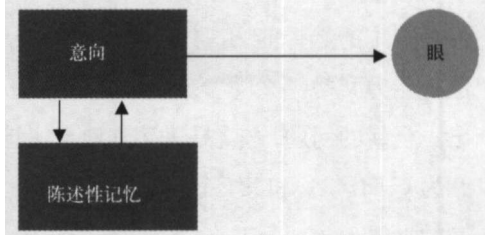


图 18

① 把意识归约为“看见”是一简化。虽然视觉信息之处理占了大脑处理信息的百分之八十五,但并不能把大脑处理信息完全等同于“看见”。因此, $C(n)$ 除了归约为  $C(n-1)$ 外,还可以归约为其它原初意识因子。这样,意识解释的递归链不只一条或若干条,实为一树状结构。

表示,而梦中看见过程可用模型图 18 表示。两者有一共同意识参数,这就是意向性(或看的意愿)。这样一来,我们得到两个比“看见” $C(n-1)$  更为原初性的意识参数,一个是意向性,我们记为  $C(n-2)$ ,另一个代表“梦中看见和醒时看见之不同”的  $C'(n-2)$ 。

在哲学上,意向性一直是最重要的内省概念,整个现象学就建立在这一原初性意识概念之上。因此,现象学方法就可以和系统论模型整合起来。但是,一旦用意向性来解释“看见”,我们立即发现,意向性和陈述性记忆耦合指向外部对像一样构成了“看见”,<sup>①</sup>而“梦中看见”无客观性可言。而现象学把意向性指向外部对像视为自明的,并认为这就是客观性。从意识的递归解释看来,这是不严格的。这样现象学必须经过精确化的改造,才能成为新的综合之一部分,和意识的递归解释整合起来。至于要准确地定义“意识到客观性(外部世界的存在)”或信息开关对外开放,则需要界定另一个更原初的意识参数  $C'(n-2)$ 。

$C'(n-2)$  必须刻划“梦中看见”和“醒时看见”的不同。大多数专家都同意梦中的看见缺乏注意力和记忆力,这样作为代表不同于“梦中看见”的“醒时看见”之意识可归为“注意力加短期记忆”。我们立即发现,这恰恰是克里克对意识的定义[克里克 1997]。显然,直接把意识等同于“注意力加短期记忆”是不确切的,因为梦中也有意识。准确地讲,克里克对意识的定义实际上只是意识递归解释链中的一环。

$C'(n-2)$  能否进一步化约?人人都知道何为注意力。它包含两个要素,一是自主的选择能力,二是精神集中于被选中之对像。克里克曾 据大脑前扣带回受损病人无意志力认为人的意识处于大脑中“前扣带回”。显然,克里克又一次误把意识递归解释中的一个环节当成意识。其实,只有这一链不断延伸下去,才能解释意识。如精神集中于被选中的对像可以进一步化约为主体给对像以主观评价。而且,只有存在自主的选择能力,才能定义意志。否则意志和被强迫就没有分别了。而自主的选择能力又是甚么呢?它能否归为更原初的 B 和更原初的 C 呢?在我们面前,展现出一全新的问题,这就是去建立对自主选择能力的解释的模型。看它怎样递归地化约为  $C(n-3)$  和其它行为参数。今天,意识解释之链正在往深处延续,一直指向未知的世界。

① 当图 18 中由陈述性记忆向意向性反馈破坏时,人可以有 REM 睡眠,但没有梦境。某些大脑顶叶受损的病人就会发生这种情况。因此,意向性指向视觉记忆并不构成看见,只有和视觉记忆耦合才构成各种看见。

## 六 唯物论还是唯识论：终将消失的界线？

如果有一天人类真的找到了意识递归解释的长链，它由 F、K、……组成，对意识的科学解释总算是实现了。但是，这是一种甚么样的科学解释呢？根据迄今为止我们对科学的理解，科学解释之所以重要，一是它具有普遍（公共）的真实性，二是它刻划了研究对象和人可以控制的变量（参数）之间的关系。正因为如此，一旦达到对某一事物工作原理的科学解释，我们就有可能模拟它以至用人工将其创造出来，正如根据仿生学原理制造具有某种生物功能的仪器一样。意识无疑是大脑的功能，那么对意识的递归解释能否帮助我们制造具有意识的计算机呢？

我认为，这是大可怀疑的。让我们来看 F、K、……长链。其中第一环 F 我们已大致知道，它是一个和计算机完全不同的生物性自组织系统，也就是其组件（神经元）只能靠介质互相沟通，这和今日我们制造的机器都不相同（社会组织倒有点类似）。但是，如果有一天出现了生物计算机和 DNA 复制构成的仪器，利用这些生物有机性的可控参数，是否会有人造意识呢？我对此亦是怀疑的。

让我们再来分析 F、K、……长链，虽然其中大多数环节尚不清楚，但我们已看到，其中相当多环节是人和动物共有的，如神经网络和做梦等。特别是随着长链的不断伸展，其后面环节的 B 参数和 C 参数将越来越多是动物性的。例如对于内省变量，当它们被归约为“选择能力”、再由“选择能力”化约为对身体位置的控制时，人虽然不能通过内省知道动物是怎样行动，但动物对运动肌控制和人类类似，可以自主地选择其身体处于甚么空间位置。<sup>①</sup>事实上，F、K、……链下端很可能类似于生物（动物）控制能力的进化树！换言之，意识和人类起源一样，是动物演化长链的结果。而意识的递归解释，本质上是去追溯意识在生物世界起源的历史过程。我们看到认知理论、心理学、神经科学和进化论的合流。要知道，达尔文进化论作为人类起源的科学解释已确立了一个半世纪，但我们尚不能模拟演化。意识问题很可能也是一样。

这样一来，意识的递归解释对人类的意义，可能和达尔文的进化论类似，只是人对自身认识是革命。其认识论意义远大于制造人工智能的意义，正如达尔文进化论对人精神世界的冲击远超过由它而引发改造世界能力的扩张那样。众所周知，达尔文进化论一度造成宗教和科学的严重对立，几乎颠覆了自现代性确立以来终极

<sup>①</sup> 必须注意，该递归解释链是从梦的模型开始的，因爬行动物没有梦，该链被限定在哺乳动物内，故递归解释链不是整个生物演化链。



关怀和理性互不干扰的心灵。它所带来的精神风暴以及由引起的社会动荡席卷十九世纪下半叶和整个二十世纪,其思想对宗教之挑战至今仍然存在。既然意识的科学解释和进化论类似,它又会对人类未来的精神造成何种冲击呢?

对这一切做预测是不可能的。但是,如果意识的科学解释真是一条或若干条递归归长链,它带来的影响必定和达尔文进化论有根本的不同。达尔文进化论造成的冲击在某种意义上缘于“物竞天择”机制的单纯性以及对其价值的肯定,意识的递归解释长链中不存在类似的单纯的可以误用之事实(和机制)。它是由行为参数 B 和内省因素 C 两个序列组成。科学地解释意识,必须去追溯这两个序列。对于 B 序列,这是一系列受控实验证明的行为参数之间的关系,它对应着我们对大脑神经系统结构和功能不断深入的认识。对于 C 序列,其不断深化之追溯是一系列严密而准确的意识内省分析,意识不断地被化约为更为基本而自明的要素。

这第二个过程是哲学式的,有点像今日之现象学。但和现象学有重要不同,因为它必须把内省才能理解的意识视为更为原初性意识和行为参数之组合。在这一意义上,它更接近佛学中的唯识论的展开,但要反过来进行。当然,它并不简单地等同今日之唯识论,因为其所有环节不能仅仅思辨地进行,用原初性意识建构较复杂意识的模型如涉及可观察变量,它必须是可检验的。或者说,它以整个科学理论成就或数学式清晰的内省作为思辨展开的基础。

如前所说,意识科学解释的达成,需要递归长链中行为主义变量 B 和内省变量 C 差别的逐渐消失。用行为主义变量 B 构成的解释是唯物论的,而用内省变量 C 构成的解释是唯识论式的。换言之,在意识递归解释 C 序列追溯过程中,撇开行为主义变量,意识将变得越来越朦胧,先是醒和梦不可区分,后是和本能的自我感受互相交融。最后在“我有”之消解中,自我意识消失在茫茫空无的宇宙之中最后和自然合一。两者之间差别的逐渐消失会造成一个惊人的结果。这就是唯物论和唯识论界线的消失。

表面上看,唯物论和唯识论界线的消失这是不可思议的。前者是把意识作为物质之属性,后者把世界视为意向性所建构,两者互不相容。人们很难想象两种互相矛盾的哲学立场会融为一体。但是,如果我们进一步思索两者的差异,就发现它们并非如表面所呈现的那样对立。为甚么唯物论要在哲学中确立客观世界的存在为第一性的前提?关键是判别现象真实性之需要。即真实的现象之所以不是大脑之虚构,乃是因为它和客观存在相一致。然而,当我们思索上述唯物论关于真实性原则时,发现它只是近似成立的。

我曾在有关著作中指出,“与客观符合为真”之原则应该被“受控实验可重复性”取代。当一个现象的观察不能被受控实验重复,它和幻觉虚构是不能区别的[金观涛 1989]。进一步讲,世界能表述为客观存在是有条件的,这就是主体和它耦合时(观察世界)对它的干扰可以忽略不计。[金观涛 1992, 页 139—148]

那么,受控实验可重复的前提又是甚么呢?它首先建立在“人想重复”之上,“人想重复”在意识的递归解释中,属于内省性的 C 序列。这里 B 序列的元素和 C 序列元素的区别开始消失了!

如果上述分析成立,意识科学解释的达成会导致唯物论和唯识论界线的消失。它是科学和宗教二元分裂的终结么?如果是,则是现代性建立以来最大的改变。自苏格拉底提出“认识你自己”以来,用认知来克服死亡一直是希腊哲人的目标,随着新柏拉图主义被融入基督教,它被包含在用理性认识上帝之中。但随着现代性之形成,理性和终极关怀的二元分裂宣告用认知来克服死亡是不可能的。那么,唯物论和唯识论界线的消失意味着甚么呢?我不知道。或许,人类已站在我们完全不理解的新时代门槛上。

## 参考文献

- 艾什比 1991.《大脑设计》. 乐秀成、朱熹豪等译. 北京: 商务印书馆.
- Foerster, H. 1984. On Constructing a Reality. in Watzlawick, P. (Eds), *The Invented Reality: How Do We Know What We believe We Know. (Contribution to Constructivism)* W.W. Norton and Company, New York.
- Hobson, A. 2001. *The Dream Drugstore, Chemical Altered States of Consciousness*. The MIT Press.
- 霍布森 1999.《梦与疯狂:解读奇妙的意识状态》. 台北: 天下远见出版股份有限公司.
- 金观涛 1989. 奇异悖论——证伪主义可以被证伪吗?《自然辩证法通讯》(2).
- 金观涛 1992.《人的哲学——论“科学与理性”的基础》. 台北: 台湾商务印书馆股份有限公司.
- 金观涛 2005.《系统的哲学》. 北京: 新星出版社.
- 克里克 1997.《惊异的假说》. 台北: 天下文化.
- 拉维 2002.《睡眠的迷人世界》. 潘震译. 台北: 远流出版事业股份有限公司.
- 罗柯 2007.《我们为什么要浪费时间睡觉:梦的科学解析》. 吴妍仪译. 台北: 猫头鹰出版.
- 李葆明 2005.《神经与脑科学》. 台北: 世潮出版有限公司.
- 史奎尔、肯戴尔 2001.《透视记忆》. 洪兰译. 台北: 远流出版事业股份有限公司.