

利贝特实验是否证伪了自由意志？

——基于相容论视角的哲学考察

董泓玮

苏州大学政治与公共管理学院哲学系，江苏 苏州

收稿日期：2026年3月6日；录用日期：2026年3月25日；发布日期：2026年4月8日

摘要

本杰明·利贝特(Benjamin Libet)于20世纪80年代开展的经典神经科学实验，揭示了大脑在主体主观意识到行动意图之前，就已经产生了无意识的神经准备活动(准备电位)。这一实证发现对建立在传统笛卡尔二元论基础上的自由意志观念提出了前所未有的严峻挑战。本文旨在梳理利贝特实验的理论背景、深远影响及其在后续研究中受到的多维学术批评，并在此基础上，提出一种基于相容论(Compatibilism)视角的系统性重构与反驳。本文详细论证指出，利贝特实验的悲观结论实际上建立在特定的、有待商榷的形而上学预设之上，且其在实验范式设计上未能有效区分无理由的随机动作与基于规范性理由的理性决策。通过引入行动哲学中的意图层级结构理论(区分远端意图与近端运动指令)、认识论中关于“理由空间”与“因果空间”的严格界分以及心灵哲学中基于复杂系统理论的涌现论视角，本文系统地阐述了一种自然主义的相容论自由意志观，并最终澄清：利贝特实验实质上解决的是“无理由自发动作的微观神经启动机制”问题，而非宏观层面“自由意志是否存在”的本体论问题。

关键词

利贝特实验，自由意志，相容论，理由空间，涌现论

Does the Libet Experiment Falsify Free Will?

—A Philosophical Investigation from a Compatibilist Perspective

Hongwei Dong

Department of Philosophy, School of Politics and Public Administration, Soochow University, Suzhou Jiangsu

Received: March 6, 2026; accepted: March 25, 2026; published: April 8, 2026

Abstract

Benjamin Libet's classic neuroscience experiments in the 1980s revealed that the brain generates

文章引用：董泓玮. 利贝特实验是否证伪了自由意志? [J]. 哲学进展, 2026, 15(4): 230-237.

DOI: 10.12677/acpp.2026.154156

unconscious neural preparatory activity (readiness potential) before the subject becomes consciously aware of the intention to act. This empirical finding posed an unprecedented challenge to traditional concepts of free will based on Cartesian dualism. This paper aims to review the theoretical background, profound impact, and multidimensional academic criticisms of the Libet experiment, and on this basis, proposes a systematic reconstruction and rebuttal from a compatibilist perspective. The paper argues in detail that the pessimistic conclusions of the Libet experiment rest on specific, debatable metaphysical presuppositions, and its experimental paradigm fails to effectively distinguish between unreasoned random actions and normative reason-based rational decisions. By introducing the hierarchical structure of intentions in the philosophy of action (distinguishing distal intentions from proximal motor commands), the strict epistemological distinction between the “space of reasons” and the “space of causes”, and the emergentist perspective based on complex systems theory in the philosophy of mind, this paper systematically elaborates a naturalistic compatibilist view of free will, clarifying that the Libet experiment essentially resolves the “micro-neural initiation mechanism of unreasoned spontaneous actions”, rather than the ontological question of “whether free will exists” at the macro level.

Keywords

Libet Experiment, Free Will, Compatibilism, Space of Reasons, Emergentism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 利贝特实验的问题背景、实验影响与面临的挑战

自由意志(Free Will)问题不仅关乎人类是否具备自主选择行动的能力,更与人类道德责任、法律惩罚以及社会伦理秩序的基础紧密相关。在哲学史中,自由意志问题总与其背后的形而上学预设有关。决定论(Determinism)主张宇宙万物(包括人类行为)均由先前的物理状态和自然规律严格决定;自由意志论(Libertarianism)则坚持人类拥有超越物理因果链条的绝对选择能力;而相容论(Compatibilism)则试图在决定论的物理世界中为自由意志寻找生存空间。然而,进入20世纪下半叶,随着认知神经科学、脑成像技术以及电生理学的高速发展,自由意志问题也经历了问题的范式转换——从扶手椅上的哲学思辨,转向了实验室里的实证科学检验。在这一历史进程中,美国加州大学旧金山分校的生理学家本杰明·利贝特(Benjamin Libet)于20世纪80年代开展的一系列开创性脑电图(EEG)实验,无疑具有里程碑式的意义。这些实验不仅对传统的自由意志观念提出了量化的实证挑战,更标志着“神经哲学”(Neurophilosophy)这一交叉学科在自由意志研究领域的正式确立。

本文的核心目标并非简单地复述或总结现有对利贝特实验的批评,而是在现有文献的基础上,提出一种具有整合性与建构性的相容论回应。现有研究或聚焦于实验的方法论缺陷(如Dennett),或进行概念辨析(如Mele),或提出新的神经科学模型(如Schurger),但往往未能将这些不同层面的批评整合到一个统一的、能够积极捍卫自由意志的理论框架中。本文创新地将行动哲学中的意图层级理论、认识论中的“理由空间”与“因果空间”的严格界分,以及心灵哲学中基于复杂系统理论的涌现论视角,这三种理论资源系统地结合起来。通过揭示它们之间的内在关联,本文试图论证:利贝特实验的悲观结论源于对自由意志概念的过度简化,而上述三种理论的整合恰好能够提供一个层次分明、解释力更强的相容论自由意志图景,从而解决单一理论视角难以完全回应的“意识滞后是否等同于意志缺失”这一核心困境。

1.1. 准备电位的发现与利贝特实验的经典范式

要深入理解利贝特实验的意义,必须追溯其科学前史。1965年,德国神经科学家汉斯·科恩胡伯(Hans Kornhuber)和吕德尔·德克(Lüder Deecke)在研究人类自主运动时取得了一项重大发现。他们通过头皮脑电图记录受试者在进行自发运动(如按键或弯曲手指)时的脑电活动,发现大脑运动皮层及辅助运动区(SMA)在肌肉实际收缩前,会出现一种缓慢的、逐渐上升的负向电位变化。这种电位被命名为“准备电位”(Readiness Potential,简称RP)。研究表明,RP通常在动作发生前约800至1000毫秒(甚至更早)就开始出现。这一发现震惊了当时的科学界,因为它无可辩驳地表明:大脑在执行一个看似瞬间的自主运动之前,实际上已经经历了较长时间的无意识神经准备阶段。

利贝特敏锐地捕捉到了准备电位发现背后的哲学意蕴。他的核心研究动机是:精确测量主观的“意识意图”(conscious intention)与客观的“准备电位”(RP)之间的时序关系。如果意识意图是动作的真正原因,那么在时间顺序上,意识意图理应先于大脑的神经准备活动;反之,如果神经准备活动先于意识意图,那么传统的自由意志观念将面临颠覆。为了实现这一测量,利贝特设计了一个极为精巧的实验范式。受试者被安排坐在椅子上,头皮上连接着脑电图(EEG)电极以记录准备电位,手臂上连接着肌电图(EMG)电极以记录肌肉收缩的准确时间。受试者被要求注视前方屏幕上一个快速旋转的示波器光点(该光点每2.56秒旋转一圈,类似于一个极快的钟表秒针)。实验指导语要求受试者在他们“感到有冲动或意愿”(urge or intention)去弯曲手腕或手指的任何时刻,完全自发地做出动作,不要预先计划。受试者需要记住他们首次明确意识到按键意图时光点所在的位置,并在动作完成后报告该位置。这个主观报告的时间点被称为W时间(Will time,意图时间)。同时,仪器客观记录了肌肉开始活动的M时间(Movement time)以及准备电位开始出现的RP时间[1]。

1.2. 实验结果的哲学冲击与“意识意志错觉”假说

利贝特实验的统计结果显示:若将肌电图记录到的肌肉开始活动的时间(M时间)设定为0毫秒,那么大脑准备电位(RP)的起始时间平均在动作前约550毫秒(-550 ms);然而,受试者主观报告的意识意图出现时间(W时间)平均仅在动作前约200毫秒(-200 ms)。这一时间差(约350毫秒)构成了利贝特实验的核心发现:大脑的无意识神经活动(RP)显著地先于主体对行动意图的主观觉知(W时间)。

基于这一坚实的实证数据,利贝特及其追随者推论出:自主行为的启动实际上是由大脑无意识的神经过程决定的,而我们引以为傲的“意识意图”并非动作的初始因果力量,它仅仅是大脑无意识活动在后期产生的一种伴随现象(epiphenomenon)。换言之,当我们以为自己正在“自由地”做出决定时,我们的大脑其实在三分之一秒前就已经替我们做好了决定。哈佛大学心理学家丹尼尔·韦格纳(Daniel Wegner)在其《意识意志的错觉》(The Illusion of Conscious Will)中进一步深化了这一观点,他提出人类的主观能动性仅仅是一种事后建构的心理错觉。大脑的无意识机制同时产生了实际的动作和随后对动作的意识体验,由于意识体验在时间上紧挨着动作发生,我们便错误地推断是意识导致了动作[2]。神经科学家萨姆·哈里斯(Sam Harris)也据此宣称,自由意志不过是一种幻觉,人类本质上是由生化算法驱动的生化机器。

1.3. 针对利贝特实验的方法论与实证批评

尽管利贝特实验的影响力巨大,但自其发表之日起,便受到了来自哲学、心理学和神经科学内部的严厉审视与多维批评。首先是方法论层面的深刻质疑。W时间(意识意图时间)的测量高度依赖受试者的主观内省与跨模态注意力转移。受试者需要在关注内部隐秘心理状态(意图的涌现)的同时,将注意力迅速转移到外部快速移动的视觉刺激(旋转的光点)上。心理物理学的丰富研究表明,这种跨模态的注意力分配和时间标记会产生显著的系统性误差。作家丹尼尔·丹尼特(Daniel Dennett)在《自由的演化》(Freedom

Evolves)中尖锐地指出,利贝特的实验预设了一个“笛卡尔剧场”(Cartesian Theater)——即假设大脑中存在一个特定的中心位置和绝对的瞬间,意识在此时此地突然“亮起”。然而,现代认知科学表明,意识经验在时间和空间上都是弥散的、分布式的处理过程。将主观报告的模糊心理时间与毫秒级精度的客观神经电位时间直接对齐,存在不可逾越的测量学与认识论鸿沟[3]。

此外,实证层面的后续研究也对利贝特的结论提出了挑战。Trevena 和 Miller (2002)的脑电图研究表明,即使在受试者被要求关注意图但最终决定“不”做出实际动作的情况下,大脑依然会出现类似准备电位(RP)的负向脑电活动。这一关键发现极大地削弱了 RP 作为动作“必然前兆”或“无意识决定”的地位,暗示 RP 可能仅仅反映了对任务的普遍关注或运动准备状态,而非不可逆转的动作指令。

1.4. 概念层面的混淆与神经科学的新解释

在概念辨析层面,哲学家阿尔弗雷德·米尔(Alfred Mele)对利贝特实验提出了极具杀伤力的批评。米尔指出,利贝特在实验设计与结论推导中,严重混淆了“冲动”(urge)、“愿望”(desire)、“意图”(intention)与“决定”(decision)等截然不同的行动哲学概念[4]。准备电位(RP)在-550 毫秒的出现,可能仅仅反映了大脑中产生了一种微弱的动作冲动或一般性的运动预期,但这绝不等同于主体做出了最终的行动决定。在人类的日常生活中,我们每时每刻都会产生无数无意识的冲动(例如想挠痒、想喝水、想打断别人说话),但并非所有冲动都会转化为实际行动。将 RP 等同于动作的决定性原因,在概念上存在严重的过度延伸与逻辑跳跃。

更直接的批评来自神经科学内部理论的最新发展。2012 年,神经科学家亚伦·舒尔格(Aaron Schurger)及其同事提出了一种关于准备电位(RP)本质的全新解释模型——“随机累加器模型”(Stochastic Accumulator Model, 或称 Leaky Competing Accumulator Model)。该模型指出,在利贝特实验这种缺乏外部明确线索、要求受试者在漫长时间段内随意自发行动的特殊情境中,大脑的自发神经背景噪声(spontaneous neural noise)会处于不断的随机波动之中。当这种随机噪声的累积偶然跨越了某个特定的运动执行阈值时,就会触发实际的动作。舒尔格团队通过计算机模拟证明,所谓的准备电位(RP),实际上只是在动作发生前对这些随机神经波动进行逆向时间平均(time-locked averaging)所产生的统计学错觉[5]。这意味着,大脑在-550 毫秒时并没有做出任何特定的“无意识决定”,RP 仅仅反映了神经系统的自发波动逐渐接近运动阈值的动态物理过程。这一革命性的模型从根本上动摇了利贝特实验将 RP 视为“无意识决定”的神经学基础,使得基于 RP 否定自由意志的论证失去了最核心的实证支撑。

2. 基于相容论视角的考察与批评

尽管上述批评从实证数据、测量方法、概念界定以及神经科学新模型等多个维度深刻指出了利贝特实验的局限性,但要系统、彻底地回应其对自由意志的挑战,仅停留在指出实验漏洞的破除工作是不够的。我们还需要构建一个能够整合现代神经科学数据,同时又能捍卫人类主体能动性的积极理论框架。本文认为,利贝特实验的悲观推论过程,实际上隐蔽地预设了一种严格的笛卡尔二元论(Cartesian Dualism)和还原论(Reductionism),并将自由意志狭隘地等同于脱离物理因果链条的绝对初始原因(即神学意义上的“第一因”)。基于现代相容论(Compatibilism)的视角,本文主张:自由意志并非微观神经元放电的无因果起点,而是宏观认知系统基于规范性理由进行复杂审思、长期规划与自我调节的自然能力。

2.1. 意图的层级结构与控制权的动态下放

行动哲学(Philosophy of Action)的深入研究表明,人类的意图并非单一维度的心理状态,而是呈现出高度复杂的层级结构。迈克尔·布拉特曼(Michael Bratman)的规划理论(Planning Theory of Intention)以及伊丽莎白·帕切里(Elisabeth Pacherie)的意图级联模型(Cascade of Intentions)指出,人类的能动性建立在不

同时间尺度和抽象层级的意图网络之上[6]。我们可以将其大致划分为三个层级：“远端意图”(distal intention, 或称未来导向意图)、“近端意图”(proximal intention, 或称当前导向意图)以及“运动意图”(motor intention)。远端意图是指向未来的、经过深思熟虑的宏观计划(例如“我决定明天下午两点去参加心理学实验”);近端意图是将宏观计划锚定在当前情境中的指令(例如“我现在坐在椅子上,准备按照指导语按键”);而运动意图则是即刻执行具体动作的微观神经指令(例如“在这一毫秒收缩右手的特定肌肉群”)。

从上述层级模型的视角来看,利贝特实验所考察的,仅仅是最低层级的运动意图,即在数百毫秒内执行一个极其简单的无意义动作的冲动。然而,在整个实验情境中,受试者绝对不是处于完全无意识的被动状态。受试者首先形成了参与实验、理解并同意遵循指导语、并在规定时间内自发按键的远端意图与近端意图。这一宏观决策过程是由受试者的意识深度参与、理性权衡并主导的。当远端意图确立之后,大脑的高级执行系统便将具体的微观操作任务交给了底层的自动神经机制。受试者实质上是将“具体在哪个随机的毫秒按下按键”的控制权,动态地下放给了无意识的运动皮层。

这种控制权的动态下放不仅不是自由意志的丧失,反而是人类认知系统高效运作的必要条件和演化优势。在驾驶汽车、演奏钢琴、进行高强度体育运动甚至日常的流利对话时,主体确立了宏观目标和战略方向,而具体的肌肉协调、句法生成则完全交由无意识的专用程序高速完成。近端动作的无意识启动,正是远端有意识决策的忠实执行结果。利贝特实验以微观层面被刻意下放控制权的近端动作,来试图推翻宏观层面有意识的远端决策的自主性,在逻辑上犯了严重的范畴误用(category mistake)。自由意志的核心价值主要体现在远端意图的制定、维持与修正上,而非微观运动指令的即时生成上。

2.2. 理由空间与因果空间的认识论界分

威尔弗里德·塞拉斯(Wilfrid Sellars)在其名篇《经验主义与心灵哲学》中,提出了“理由空间”(space of reasons)与“因果空间”(space of causes)的著名区分[7]。约翰·麦克道威尔(John McDowell)等当代学者进一步深化了这一理论。这一区分对于理解自由意志至关重要。自由意志的核心特征在于,主体能够依据规范性的“理由”(reasons)进行选择 and 行动,而非单纯受盲目的物理“原因”(causes)的因果驱使。理性的审思发生在主体需要权衡不同选项的道德价值、长远利益、社会规范与个人信念的复杂情境中。

当我们审视利贝特实验的范式设计时,会发现它刻意剥离了所有规范性的理由。实验指导语要求受试者“不要预先计划”,仅凭“自发的冲动”行动。在这种极端的人为设定下,何时弯曲手腕对受试者而言缺乏任何实质性的理由或价值差异。在第3秒按键与在第8秒按键,不会对受试者的生活、道德评价或实验报酬产生任何不同的影响。这在哲学上构成了一个典型的“布里丹之驴”(Buridan's ass)困境:在面对完全等效、毫无价值差异的选项时,理性审思机制无法提供决策依据。在排除了理由权衡的无差异情境中,大脑的理性决策系统处于“挂机”状态,只能依赖底层的自发神经波动(正如Schurger模型所述的背景噪声)来打破对称性,从而触发动作。

该实验确实证明了,人类在完全无理由的状态下,会依赖无意识的神经机制来启动动作,但这完全属于因果空间的物理机制运作。然而,真正的自由意志的运作属于理由空间。当人类面对涉及职业选择、伴侣挑选、道德判断或长期人生规划的复杂抉择时,其决策过程充满了意识的深度参与、逻辑的严密推理、情感的权衡和价值的考量。将一种刻意剥离了理由、审思和意义的实验范式所得到的微观结果,强行推广至所有包含理由和审思的宏观人类行为,在认识论上是缺乏合法性的。

2.3. 复杂系统、涌现论与非还原物理主义的自主性

利贝特实验的悲观推论,在深层本体论上隐含了一种还原论(Reductionism)和二元论(Dualism)的混合假设:如果神经活动先于意识觉知,则意味着是“大脑的物理机制”代替“主体”做出了决定。这种表述

将大脑的物理过程与主体性对立起来，仿佛主体是一个独立于大脑的事物。无法与大脑的运作同时存在。

相容论者，特别是支持非还原物理主义(Non-reductive Physicalism)和复杂系统理论(Complex Systems Theory)的学者主张，自由意志应被视为大脑 - 身体 - 环境系统在宏观层面的涌现属性(emergent property)。艾丽西亚·华雷罗(Alicia Juarrero)等人的研究表明，宏观层面的系统属性不能简单还原为微观层面的物理部件的线性叠加。主体的记忆、性格、信念、价值观以及长期的理性训练，均以复杂的神经突触连接权重和宏观网络拓扑结构的形式存在于大脑之中。无意识的神经活动并非外在于主体的独立或异己力量，它们正是主体认知系统的物质载体和基础组成部分[8]。

当运动皮层产生准备电位时，这正是“主体”的神经系统在为行动进行准备。只要一个行为源于主体内部的认知与动机网络(包括有意识的信念和无意识的习惯)，并且未受到外部的强制(如持枪胁迫)或严重的内部病理干扰(如脑部肿瘤导致的强迫症或妥瑞氏症)，该行为即可被视为自主的。神经科学揭示了决策的微观物理基础，但这并不等同于消解了基于该物理基础涌现出的宏观主体自主性。宏观的意图和理由可以通过“向下因果作用”(downward causation)或结构性约束，塑造微观神经元的放电模式。将“大脑的决定”与“我的决定”对立起来，是混淆了不同解释层次的范畴错误。大脑的决定，在正常运作的条件下，正是我的决定。

3. 上述观点的潜在理论批评与对利贝特实验的最终回应

本文提出的基于相容论的系统性重构，在理论上不可避免地会面临来自不同哲学阵营的批评。为了使论证更加严密，本节将探讨两种主要的潜在批评，并给出相应的理论回应，最后明确界定利贝特实验在自由意志探讨中的真实理论位置。

3.1. 应对强决定论的挑战：理由回应与法兰克福范式

首先是来自不相容论者，特别是强决定论者(Hard Determinists)可能提出的最强反驳。他们可能指出，即使本文成功区分了远端意图与运动意图，并强调了理由空间在宏观决策中的作用，但这一切本身依然处于决定论的链条之中。远端意图的形成、理性审思的过程，甚至主体对理由的权衡和回应能力，最终都可以追溯到主体无法控制的先前的物理状态(如基因遗传、早期家庭环境、社会文化背景)和自然规律。因此，决定论者会认为，相容论所描绘的自由意志只是一种表面现象，主体依然无法逃脱因果链条的决定。如果主体对理由的“回应”本身是被决定的，那么这种回应还能被称为“自由”的吗？这是相容论面临的来自强决定论的最有力的挑战。

对此，可以做出以下回应：相容论并不否认物理世界的因果决定性，而是主张对“自由”的概念进行更为务实和符合人类生活形式的重新界定。在现代相容论框架下，特别是哈里·法兰克福(Harry Frankfurt)的层级动机理论(强调二阶意愿对一阶欲望的认同)或约翰·马丁·费希尔(John Martin Fischer)的“理由回应”(reasons-responsiveness)理论中，自由并不意味着行为缺乏因果前件(即要求非决定论的奇迹)，而是指行为的因果链条必须经过主体内部特定的理性审思机制。只要主体的决策系统能够正常接收外部信息、理解规范性理由、根据自身价值观进行权衡，并据此控制行为，主体就是在道德和实践意义上自由的。外部初始条件确实决定了主体的神经结构，但正是这种特定的神经结构赋予了主体处理理由和适应复杂环境的卓越能力。在相容论看来，自由意志的真正对立面是外部强制、洗脑、成瘾和严重的精神病理状态，而非普遍的因果决定律。

3.2. 应对意识副现象论的挑战：意识的演化成本与因果效力

其次是关于意识副现象论(Epiphenomenalism)的残留批评。部分受利贝特实验影响的学者可能质疑，如果近端动作的启动确实是无意识的，我们如何确保远端意图和理由权衡过程中的意识体验不是一种

事后合理化(post-hoc rationalization)? 即意识可能仅仅是伴随复杂神经计算产生的一种无用的副产品, 就像火车引擎冒出的蒸汽, 对火车的运行毫无因果作用。

首先, 从演化生物学的视角来看, 意识副现象论难以解释意识高昂的“演化成本”(Evolutionary Cost)。大脑是人类消耗代谢能量最多的器官, 而支持意识觉知、工作记忆和高级执行功能的额顶叶皮层网络更是极其“耗能”。如果意识仅仅是一种毫无因果效力的副现象或事后解释器, 自然选择机制绝不会允许如此昂贵且无用的生物学特征在漫长的演化史中被保留甚至不断强化。意识的存在及其复杂的神经基础本身, 就强烈暗示了它在提升生物适应性、解决复杂生存问题上具有不可替代的演化功能。

其次, 现有的认知科学证据充分揭示了意识在高级认知过程中的真实“因果效力”(Causal Efficacy)。尽管无意识过程足以高效应对高度自动化的、习惯性的常规动作(如利贝特实验中简单的、无意义的屈指动作), 但在面对高度复杂的新异环境时, 意识的参与是不可或缺的。心理学家罗伊·鲍迈斯特(Roy Baumeister)的综合研究指出, 意识思考并非用于直接微观地控制即时的肌肉运动, 而是用于指导人类的长远行为。意识使得我们能够进行多步逻辑推理、进行心理模拟(mental simulation, 即在脑海中“试错”以避免在现实中付出致命代价)、掌握新技能、进行复杂的语言交流, 以及在必要时抑制强烈的本能冲动[9]。换言之, 意识的因果效力体现在它能够无意识地自动处理系统设定目标、重塑习惯并提供宏观的“实施意向”(implementation intentions)。

最后, 必须澄清“近端意图”与“远端意图”在行动控制中的不同分工。阿尔弗雷德·米尔(Alfred Mele)等相容论者指出, 利贝特实验的致命局限在于其仅仅考察了缺乏道德权衡和长远规划的“近端无意义动作”。在这种特定情境下, 意识将执行权下放给无意识的“自动驾驶模式”恰恰是认知资源最优分配的体现。然而, 这绝不意味着在涉及重大人生抉择(如职业规划、道德抉择)的“远端意图”形成过程中, 意识也是滞后或无效的。在这些宏观决策中, 意识通过深思熟虑设定了行为的边界条件, 随后的无意识神经活动只是在这些既定框架内执行具体的微观动作。

综上所述, 将意识在简单、无意义动作中的滞后性, 草率地过度推理为意识在所有高级认知过程中的无效性, 是缺乏充分经验支持和演化逻辑的。意识并非无用的蒸汽, 而是列车的导航系统; 它或许不直接转动车轮, 但决定了列车行驶的最终方向。因此, 基于相容论的立场, 自由意志并不要求意识微观地、即时地控制每一个神经脉冲, 只要意识在宏观的理由回应和远端意图的设定中发挥了真实的因果作用, 自由意志的根基就依然稳固。

3.3. 理论澄清：利贝特实验究竟解决了什么问题？

基于上述详尽的分析与反驳, 本文对利贝特实验的最终理论回应如下: 利贝特实验是一项伟大的神经科学探索, 但它并未触及自由意志的核心运作机制。该实验的范式设计, 通过刻意排除外部理由和内部偏好, 成功地测量了在“无差异情境”下, 大脑如何利用自发的神经背景波动来打破僵局并启动动作。因此, 利贝特实验实质上解决的是“无理由自发动作的微观神经启动机制”问题。它无可辩驳地证明了在特定的人为条件下, 无意识的神经准备可以先于意识觉知。然而, 它完全未能证明, 人类在进行基于规范性理由的、涉及长期利益和道德考量的复杂决策时, 缺乏宏观的自主性。

4. 结语

综上所述, 本杰明·利贝特的经典实验是探究人类行为神经机制的重要里程碑。它以精确的实证数据揭示了无意识神经活动在简单自发动作中的先导作用, 打破了传统哲学观念中意识对行为控制的绝对垄断地位, 促使学术界重新审视意识与行为之间错综复杂的时间动力学关系。然而, 将该实验的局部结果直接等同于“自由意志不存在”的宏大结论, 存在着理论上的过度推演与严重的范畴误用。

本文通过系统梳理现有批评，并通过引入相容论的分析框架指出，利贝特实验的根本局限在于其聚焦于剥离了规范性理由的近端微观动作，并在深层逻辑中隐含了过时的笛卡尔二元论预设。通过创新地将意图层级理论、理由空间与因果空间的界分、以及复杂系统涌现论这三种理论资源进行系统性整合，本文揭示了三者之间的内在关联：意图层级理论提供了自由意志运作的纵向时间结构，理由空间与因果空间的界分界定了自由意志运作的规范性质，而复杂系统涌现论则奠定了自由意志在本体论上的自然主义基础。三者共同构成了一个层次分明、解释力强的相容论自由意志图景，从而解决了单一理论视角难以完全回应的核心困境，即如何在承认神经科学发现的同时，捍卫人类主体能动性的真实性。而本文认为，基于理由的理性决策过程足以支撑一种自然主义的、规范的自由意志概念。面对决定论和副现象论的潜在理论挑战，现代相容论提供了一种将前沿神经科学事实与人类主体能动性相协调的坚实理论路径。

最终，利贝特实验并未否定自由意志，而是促使我们将自由意志从一种抽象的、超自然的形而上学概念，转化为一种建立在极其复杂的神经计算基础上的宏观认知能力。未来的神经哲学与认知科学研究，应当超越简单的“自由意志存在与否”的二元对立之争，致力于揭示这种宏观认知能力在不同社会情境、不同脑区网络中的具体运作机制、演化历程及其病理学边界。只有在跨学科的深度对话中，我们才能真正理解人类作为理性行动者的本质。

参考文献

- [1] Libet, B., Gleason, C.A., Wright, E.W. and Pearl, D.K. (1983) Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). *Brain*, **106**, 623-642. <https://doi.org/10.1093/brain/106.3.623>
- [2] Wegner, D.M. (2002) *The Illusion of Conscious Will*. The MIT Press. <https://doi.org/10.7551/mitpress/3650.001.0001>
- [3] Dennett, D.C. (2003) *Freedom Evolves*. Viking Penguin.
- [4] Mele, A.R. (2009) *Effective Intentions*. Oxford University Press.
- [5] Schurger, A., Sitt, J.D. and Dehaene, S. (2012) An Accumulator Model for Spontaneous Neural Activity Prior to Self-Initiated Movement. *Proceedings of the National Academy of Sciences*, **109**, E2904-E2913. <https://doi.org/10.1073/pnas.1210467109>
- [6] Bratman, M.E. (1987) *Intention, Plans, and Practical Reason*. Harvard University Press.
- [7] Sellars, W. (1956) Empiricism and the Philosophy of Mind. *Minnesota Studies in the Philosophy of Science*, **1**, 253-329.
- [8] Juarrero, A. (1999) *Dynamics in Action: Intentional Behavior as a Complex System*. MIT Press.
- [9] Baumeister, R.F., Masicampo, E.J. and Vohs, K.D. (2011) Do Conscious Thoughts Cause Behavior? *Annual Review of Psychology*, **62**, 331-361. <https://doi.org/10.1146/annurev.psych.093008.131126>