

心灵研究的“归属论”走向： 解释主义及其最新发展

高新民 张蔚琳

〔摘要〕 解释主义不直接解剖、探究心灵本身，而把主要精力集中在人的言语活动和解释实践之上，创造性地提出：心灵既不像各种形式的实在论所说的那样是实在的存在或属性，也不像各种非实在论或取消论所说的人为的虚构，而认为心灵是我们的解释性投射的产物。解释主义是有韧性的。否定和批评不但没有使之消失，反倒使之在回应中上了新台阶。在解释主义的后来发展中，不仅戴维森和丹尼特对自己创立的解释主义有改进，而且有些人对之作出了新的辩护、论证和发展，还有一些人在阐释的基础上建构了解释主义的新的理论形态。这些为我们深入探讨心灵的本质提供了新的思路和资源。

〔关键词〕 心灵；解释；新解释主义；归属论

〔中图分类号〕 B54 〔文献标识码〕 A 〔文章编号〕 1000-4769 (2019) 05-0156-10

在面对心灵究竟是什么、在自然界有无存在地位这类问题时，解释主义一反传统的思维方式，不直接解剖、探究心灵本身，不直接思考它有什么结构和功能，其物理机制和基础是什么，与物理过程是什么关系，在自然界中有什么地位，而把主要精力集中在人的言语活动和解释实践之上，着力回答理解和解释人的言语行为如何可能以及心理状态为何、如何“具有”的问题，经过对解释条件、根据的丝丝入扣的探讨，最终形成了一种全新的、令人震惊的心灵哲学理论。其基本观点是：心灵既不像各种形式的实在论所说的那样是实在的存在或属性，也不像各种非实在论或取消论所说的人为的虚构；人本无心灵，本无意向状态，它们是我们的解释性投射的产物，换言之，

是我们为了解释人的行为而强加给人或归属给 (attribute to) 人的；精神或心灵不是像自古以来人们天经地义地认识的那样，是实在进化出来的，而是人为解释的需要而设定的。对此，丹尼特做了精辟的概括：“人的心灵本身是人们在重构人脑时为了方便而创造出来的一种人工制品”。^①按照丹尼特 (D. Dennett) 的划分，经典的解释主义有投射主义 (projectivism) 和规范主义 (normativism) 两种形式。前者由戴维森所创立，后者的代表人物是他自己。^②尽管它们极富创意，开拓了思考心灵哲学问题的新思路，但“和者甚寡”，被挑出的毛病很多，如违反常识和直觉，陷入了副现象论，等等。解释主义是有韧性的。否定和批评不但没有使之消失，反倒使之在回应

〔基金项目〕 国家社会科学基金重大项目“东西方心灵哲学及其比较研究” (12&ZD120)

〔作者简介〕 高新民，华中师范大学心灵与认知研究中心教授、博士生导师；

张蔚琳，华中师范大学心灵与认知研究中心博士研究生，湖北 武汉 430079。

中上了新台阶。在解释主义的后来发展中，不仅戴维森和丹尼特对自己创立的解释主义有改进，而且有些人对之做出了新的辩护、论证和发展，还有一些人在阐释的基础上建构出了解释主义新的理论形态。这里，我们将重点探讨它后来的发展，特别是这些新的形态及其辩护。

一、解释主义的不同解读与分类

解释主义无疑在解答心灵的本质问题时提出了一种革命性、建设性的思想。它引出新的有意义的但过去被掩盖、遮蔽的问题是，人们关于心理现象的认识究竟是实在的反映，还是归属或解释的产物？换言之，心灵究竟是真实存在的，还是我们为了解释人的行为而加之于人的？在现实生活中，即使不说一切心理现象都是归属、投射的产物，但至少在许多情况下可看到这样的事实，即对象本没有心理现象，但为了解释的需要，我们真的把心理现象归属于他（它）们了。例如，在解释婴幼儿的许多让人忍俊不禁的行为时，我们常常会说他们有某某信念、愿望、想法等。很显然，他们不可能有这些东西。另外，在解释会话机器人、服侍机器人等有所谓智能的人工产品时，我们常常说他们有某某情感、想法、意图等。这些显然是解释的强加。因为解释主义有这样的解释力和合理性，因此一经诞生就赢得了许多粉丝。当然也应看到，由于它建立在关于合理性和无矛盾性的假定之上，因而又受到了许多质疑，如：（1）认知科学和神经科学已发现了大量还原的证据，因此解释主义否定还原便面临着与科学的矛盾。（2）解释者对心理状态的归属可能成为解释者的任意投身，如果是这样，怎样判断一种归属比另一归属更好。（3）意识是心理王国的重要组成部分，似乎不受不确定性的影响，他人的意识肯定不是我们在我们自己身上所经验的东西，因此不是解释主义所说的归属的产物。如此等等。正是这些问题，成了解释主义向前发展的动力源泉。

解释主义在当代发展的一个重要表现是，赞成或倾向解释主义的人纷纷提出自己的新的解读。解释主义面临的经常性的问题是，怎样理解、说明心理事件和物理事件？

一般认为，不外两种方式。一是用谓词来说明，比如说：说一事件是物理的，等于说有一个物理描述，它足以正确地说明这个事件。换言之，

如果用物理谓词去描述，一事件即为物理事件。同样，如果用心理谓词去说明，那么它就是心理事件。二是根据属性例示来说明，如说：如果一事件是物理属性的例示，那么它就是物理事件。同样，如果它是心理属性的例示，即为心理事件。很显然，戴维森用的是第一种说明方式。最近，有学者尝试用后一方式来加以说明。根据这一模式，心理事件同时有两种属性，一是心理属性，二是物理属性，即任何心理事件同时可成为两种属性的例示。如果是这样，那么异常一元论中的一元论就可看作是一种个例同一论，因为它不外是主张：每一心理事件实即一物理事件。当然，不能倒过来说，每一物理事件都是一心理事件，因为有纯粹的物理事件，如飓风。另外，物理事件的特殊性还在于：它可以例示一个以上的属性，如一张桌子既可例示颜色、重量等属性，还可例示形体等属性。

由于有这样的观点，被重新阐释的异常一元论便能断言：一物理事件可同时例示这样的心理和物理属性，它既是一中枢现象，又能表现为诸如相信天要下雨这样的信念。换言之，这两个属性可同时属于同一个事件。这种一元论还承认：所有心理事件都能例示物理属性。这一观点强调的是：每一心理事件作为事件是可同一于物理事件的。

如此被重新阐释的异常一元论坚持这样的异常论，即否认存在着控制人的思想和行动的严格的、合规律的规则。它包括两方面的内容。一是关于心理的异常论，它否认有把心理状态与心理状态关联起来的规律，进而不承认有把FP整合到更大理论中的可能性。二是心理物理论，它否认有把心理事件与物理事件关联起来的规律，而且不承认有能够把FP还原为物理学的桥梁原则。

很显然，这一理论是特定意义上的物理主义，因为它承认心理事件还可还原式地同一于物理事件，即赞成个例同一论。但不赞成属性同一论。因为后者是类型同一论，即是强同一论。

由于坚持个例同一论，因此它就能承认心理有因果作用。根据个例同一论，一心理事件的因果作用是根源于与它个例同一的物理事件。如此理解，戴维森的三个原则也就能自圆其说了。根据第三个原则，只有例示了物理属性的物理事件才有因果作用，才服从严格的决定论规律。根据第一个原则，心理事件可作为原因引起物理事件。之所以如此，是因为心理事件有物理属性，或例

示了物理属性。

怎样理解严格规律？根据一种阐释，严格规律就是既能得到肯定事件确认，又能得到反事实支持的普遍原则。它不同于余者皆同的规律，后一规律是有例外的，因此会附带有余者皆同的从句，即它们只存在于一切别的方面都相同这样的条件之下。严格规律则不同。它是没有例外的，在起作用时，它是不依赖于有选择的条件的。另外，严格规律还有一个标志性特点，即它从属于基本的科学理论。所谓“基本”是指，它可作为解释别的一切自然现象的根据、基础。这也就是说，基础理论描述了一个封闭的、无所不包的系统。根据这一对严格规律的规定，不可能有严格的心理学规律，例如 FP 尽管有许多一般原则，如关于信念、愿望的原则等，它们尽管事实上制约着人们对思想和行为的解释，但在许多情况下，常常会有反例，或被违背。因此它们不是严格规律。FP 的原则能否被改造成严格的规律？根据关于严格规律的限定，即它一定是有例外的，从属于封闭的、全面的理论。FP 的原则是不可能达到这一要求的。因为 FP 不可能是既封闭又全面的，不可能还原为前述理论，首先，它不是封闭的，因为非物理事件可引起心理事件，即心可超越于物理的东西，它也不能具有广泛的包容性，因为心理事件不能包含物理事件。同样，也不可能严格的心理物理规律。

根据这一阐释，戴维森的非还原的唯物主义尽管也是一种个例物理主义，但同时又不同于一般的个例物理主义，因为后者在戴维森看来，没有抓住物理主义的基本精神，尤其是没有对心物关系作出令人满意的阐释，例如它对心理属性与物理属性的关系态度暧昧。戴维森个例物理主义的特点在于，它有自己的随附论。这种随附论要说的是心对物的依赖关系，即事物从心理上所表现出的样子恰好是从物理上表现出来的样子，因此只要能确定物理属性的分布，就能确定心理属性的分布。奥迪（R. Audi）：“说心理属性随附性物理属性，其实是这样断定了一种关系，（1）任何两个事物都不可能共有它们的物理属性，进而不能在它们的心理属性上（如果有的话）相互区别开来；（2）一个有心理属性的事物是由于有一个或更多的为它所具有的物理属性而具有那个属性的。”随附性说的是两事物或属性间的依赖、决定、非等同的关系。例如一团墨点与它的形状，再如做好事与美德属性显然是不同

的，但前者又决定了后者，或后者依赖于前者。这里的关系项可以有不同的形式，如真理、事实、谓词、属性，等等。有两种形式，一是法则学的，二是概念的。^③戴维森所说的随附性主要是后者。

还有这样的解读，即主张物理事件和状态具有心理的属性。这种属性是外在于物理载体的，亦即是对物理状态是什么没有影响的属性。就像一物体运动到另一物体左边的属性，即“靠左”的属性那样。很显然，如果这样理解异常一元论，那么它就成了副现象论。

新解释主义必须回答的一个问题是，如果心理事件依赖于解释者的解释，那么它是否有真实的存在地位？如果有，那么心理事件与物理事件究竟是什么关系？如果像戴维森有时所说的那样，一事件究竟是什么类型的事件，完全是由描述所决定的，这一来，心理事件与物理事件不就成了双重言论所说的异名同实的关系吗？不就成了一个东西吗？但有时他又强调它们有因果关系。而如前所述，事件的因果关系是以有两个独立的存在为前提条件的。这里该作怎样的决断？

对于心理事件是否真实存在的问题，解释主义内部有两种说法，一是认为，心是解释者的投射，不是真实的存在。戴维森自己也有这样的表述：“拥有全部知识的解释者对于说者的意思能知道的东西就是应学的一切，这也适用于说者所相信的东西”。^④这里所说的解释者是从零开始的解释者。他要解释说者话语的意思以及说者有什么信念，他必须学习许多必学的东西，即必须具备解释所需的一切条件。可见，说者的意义和信念都依赖于解释者的解释。有新解释主义者在此基础上提出了这样激进的观点：被解释者的思想内容依赖于解释者所做的判断。“依于判断”（judgement – dependent）这个概念是由赖特（G. Wright）和约翰斯顿（M. Johnston）分别提出来的。他们认为，我们说某人有什么概念是依赖于判断的。例如一事态 A 是道德上善良的，当且仅当公正的观察者倾向于判断 A 是道德上善良的。约翰斯顿说：“关于意义或内容的事实不可能超出我们把握那个事实的理想的倾向”。^⑤质言之，某人有某信念，是因为有解释者将此信念归属于他。用双条件句表示即是：x 相信 p，当且仅当一个有适当知识的理想解释者倾向于把此信念归属于 x。

另一不同的理解是认为，心理事件尽管是用民间心理学的语言归属于被解释者的，但它不是

子虚乌有，与描述对应有某种实在存在，这实在用物理语言描述，即为物理的东西。新解释主义者强调说，解释主义尽管承认两者的个例同一，但并不认为它们是一个东西。因为它们之间存在着不相容性或范畴差异。其具体表现是：（1）心理现象是由它的意向性决定的，可用命题态度来对之做出描述，而这种描述有内涵性。物理现象则不是这样。（2）如蒯因所强调的：意向语句的不可还原性类似于翻译的不确定性。简言之，不存在把语词刺激与意义关联起来的唯一的方法，戴维森认为，存在着与同一开放行为一致的许多心理事件。（3）心理事件与物理事件之间不存在一一对应的关系，因为心理事件只能用整体论的语词来描述。例如一个人只有当能够预言别人的信念和愿望时，才能把信念和愿望归属于某人。（4）即使物理事件也可用整体论术语描述，但两种描述完全不同，因为信念和愿望是通过预设信念和愿望的人是有理性的这一点来归属的。^⑥

麦克道尔（J. McDowell）和布兰顿（R. Brandom）的创发性解读是认为，解释主义有两个关键的主张。（1）心理学解释，即根据人的做事的理由的解释，此解释带有整体论特征。这里的“整体论”是指，一特定行为只能这样从心理学上予以解释，即把信念、动机、足以导致行为的充分理由之类的复杂状态归属于该行为主体。即使心理学解释是整体性的，不同于碰碰球式的因果解释，但戴维森、丹尼特仍坚持认为它是因果解释。（2）信念一定是精确的，因而是统一的，这两个命题的意义是：第一，如果它们是对的，那么还原论、功能主义就都是错的。因为大脑、认知系统各不相同，都有自己的特殊构成和限制，因此就不可能把用心理语言描述的状态、事件与大脑关联起来，进而就没法做出还原。第二，它的一元论一定既是异常的又是一元论的。之所以是异常的，是因为心理类型不能还原为大脑类型或认知类型。之所以是一元论，是因为它认为，从心理上描述的状态仍是物理世界的组成部分，只是它截取的是带有整体论性质的东西。第三，可描述为心灵的东西必须符合两个规范标准，即（1）合理性，（2）确切性或一致性。其隐含的意义是：心理学话语不可能系统地关联于关于认知或信念加工的话语，因此认知科学或关于心灵的科学是不可能的。

根据一种新的阐释，解释主义可理解为关于心灵的规范图景或规范理论。布鲁克（A. Brook）

指出：要说一有机体或一实在有心灵，这实在必须满足某些规范。由此所决定，心灵即便是自然系统，但也根本有别于别的自然系统，甚至不同于大脑和认知功能系统。^⑦另也有人这样加以概括：有机体要具有心灵，其行为就必须符合某些规则。

最近，有人概括说，对于解释主义有这样两种常见的解读方式，一是认识论或方法论解读。根据这种解释，解释主义可看作是一种方法论或认识论。作为一种认识论说明，它可归结为这样一些准则和原则，其作用是帮助我们相互解释。也可这样说，根据这种理解，解释主义可归结为这样的一种认识论理论，它揭示了人们具有心理状态所依赖的认识论条件。二是形而上学解读。根据这种解读，解释主义可理解为一种形而上学理论，即一种关于具有心理状态的构成性条件的理论。当然，如此被理解的解释主义只是其最强的版本。因为它断言：我相信有如此这般的事情可为我详尽地加以说明，如可理解为对如此这般事情的相信。在揭示解释主义的实质时，有的人把它归结为证实主义和行为主义。米勒指出，这是误解，因为这两种理论恰恰是解释主义要避免的陷阱。^⑧米勒认为，要把握解释主义的精神实质，首先要认清它提出和关心的专门的课题。它尽管也关心理解的条件、语言的意义、心理的本质等一般性、共同性问题，心理状态是否是自然类型的问题，心理语词是否指称自然类型的问题等，但它认为，过去人们忽视了一个至关重要的问题，即具有问题或归属问题（人因什么而具有心理状态？是自在具有，还是因解释而有），而将目光聚焦于心理本质问题之上。解释主义尽管有对心理本质问题的间接、附带的回答，但关心的主要的、核心的问题是具有问题，即人们为什么以及怎样拥有心理状态这一问题。^⑨除此之外，它还有许多独立不共的问题，解释中的理性与非理性问题、解释的根据问题，是否存在着解释所假定的意义概念，等等。由于有这些与解释密切联系在一起的问题，解释主义自然成了一种解释哲学。但如前所述，由于它是一种巧妙构建起来而同时具有学科多态性的理论，因此又是一种纯正的心理哲学理论。

在最近的解释主义的重新解读与发展中，诞生了许多新的解释主义的理论形态。从时间上说，解释主义有新老之分。我们将在后面再阐释这些新的解释主义形式。老解释主义的创始人是戴维

森、丹尼特。而他们的解释主义是方法论或认识论类型的解释主义，不是形而上学的解释主义。因为他们没有对命题态度作构成性说明。他们强调，不存在自在的心理事件，只有当物理事件从心理上描述或解释时，它们才成了心理事件。可见他们不认为心理事件是从形而上学上被构成的。^⑩如戴维森说：“异常一元论主张心理事件和状态并不是归属者向自主体的纯粹的投射，相反，它认为，心理事件像物理事件一样是真实的，并同一于物理事件，状态的归属是客观的”。^⑪不难看出，他既承认归属的客观性，也承认被归属者的心理状态有客观性，而不认为，心理状态纯是投射、归属的产物。不过，他又没有倒向二元论，而坚持异常一元论，如认为心理状态不是物理状态之外的状态，而本身就是物理状态，由于用心理语言做了描述，或从心理上作了归属才成了心理状态。

从对心灵是否采取实在论态度上说，新解释主义主要有三种形式，即纯解释主义、暴露主义、归属理论或中间型解释主义。纯解释主义过分强调解释者的观念，认为人有无心灵取决于这个观念。就是说，一个人有心理状态可这样加以完全的说明，他被解释了，或他是可被解释的。不难看出，可解释性是一个人有心理状态的充要条件。这种理论可称作纯归属主义。^⑫因为它认为，人有无心理状态完全是一个解释或归属问题。很显然，其问题是让解释主义失去了客观的限制。而暴露主义要么过于神秘，要么不是一个严格的解释主义的形式。^⑬暴露主义（revelationism）是解释主义中最靠近实在论的理论。它强调，意向模式的存在不依赖于他人的解释。换言之，尽管模式的意向性可根据解释来说明，但它的有与无不取决于解释，因为它有其客观的存在地位。它不同于别的解释主义的地方在于：不把解释当作是人具有心理状态的必要条件，不承认解释者的基础地位。总之，仅在方法论的意义上，它才被看作是解释主义。^⑭中间型解释主义或归属主义是由解释主义的重要继承人米勒（B. Mölder）所创立的一种不太激进的解释主义，基本主张是：心理状态的归属不只是解释者的一种投射，而且展示了一种意向模式。当然，可解释性又是对这一模式的充分说明中不可或缺的东西。很显然，这一理论有两面性，一方面承认被解释的人有意图的形式，另一方面又承认这形式离不开解释者的解释。就此而言，它是一种折中性的解释主义。当然由于它

强调解释对于人拥有心理状态的必要性，因此这一理论才不是完全的心理实在论，而被归于解释主义阵营。应特别注意的是，尽管它承认意向模式的存在，但又认为这种意向模式是一种“脆弱的”（fragile）的实在。这种实在的脆弱性特点，可用“赘述”（pleonastic）属性或事实这一概念来说明。他说：“仅仅在赘述的意义上，才能说存在着意向属性和事实。具有这样的意向属性，得到这样的意向事实应借助它们的归属条件来解释。一方面，归属意向模式不能独立于存在着的意向事实来回答；另一方面，由于归属受到了可得到的材料和解释原则的限制，因此归属不是任意的”。^⑮他强调：理想的归属不是任意的，而有其客观的依据，这就是被解释者身上客观表现出来的东西，其中当然包括民间心理学所说的信念、愿望之类。当然，它们的本来面目需要重新刻画。

从激进程度上看，解释主义有激进与温和之分。前述的纯解释主义是接近取消论的激进的解释主义，归属理论则是其温和的形式。由于我们过去对老解释主义、归属理论等做过较多研究，且有成果发表，因此这里不涉及这些形式，只拟对一些新的、较温和的形式做一些考释和探讨。

二、弱化的解释主义与强硬的 非还原物理主义

这种解释主义是在回应批评者的责难时发展出来的。一种观点认为，解释主义对心灵的看法过于简单化。例如相信者或被解释者表现出来的言行与内心的想法、心情是不一致的，有时根本相反，如他恨死某人，但却说爱某人，并有相应的行为表现。对于这类情况，解释主义就无法做出正确的归属。之所以如此，是因为它过于简单。另外，有时，相信者的言行是虚伪的，并没有相应心理状态发生。有时，有心理状态发生，却没有言行表现。还有这样的相信者，他有足够的能力和办法，在自己有信念时，却不让解释者知道他有信念。这些足以表明解释主义是错误的。

鉴于丹尼特和戴维森等人解释主义的诸多难题，拜恩（A. Byrne）提出了弱化（weakened）解释主义这一新形式。首先他认为，解释主义表达的是这样一个先天的命题： x 相信 p ，当且仅当存在着一个有适当知识的解释者，他倾向于把信念 p 归属给 x 。^⑯其次，他承认相信者有信念依赖于解释者的解释这一原则，但不赞成丹尼特和戴

维森对解释者的规定。前者把解释者看作是普通的人，而后者把他规定为彻底的或从零开始的解释者。拜恩认为，解释者是一种“弱化的”解释者，即“有心理内容的第三人称的解释者”。这心理内容有其基础性，足以满足哲学的要求。可见这里的“弱化”有两个表现，一是强调被解释的人有心理内容，这心理内容有其基本的存在地位。这就是说，不管被解释者身上是否有心理现象，但可肯定的是，既然解释者能对他做出第三人称解释，那么就可肯定他有心理内容。其次，解释者是从第三人称角度对被解释信念做出解释的。第三，这种新解释主义不仅不隐讳旧解释主义的错误，而且还对之提出了批评。如旧解释主义认为双条件句为被解释者相信 p 提供了必要条件，拜恩认为，这是不符合事实的，因为有这样的情况，相信者在被解释时可以设法掩盖自己，如有心理状态而表现出没有，没有心理状态却表现为有，在这种情况下，解释者的解释显然就不是被解释有或没有心理状态的必要条件。换言之，“解释者判断被解释者相信 p，对于被解释相信 p 是不充分的。”第四，对解释主义的弱化还表现在，不认为双条件句提供了必要条件，只承认它提供了充分条件。根据这一弱化，解释主义的基本原则应这样表述：x 相信 p，其条件是，下面的条件能成立，即如果存在着一个有适当知识的理想解释者，他倾向于把信念 p 归属给相信者。可见，这里的修改表现在把双条件句换成了一条条件句。这一修改足以应对反对者所提出的种种驳难，如解释者常作错误的归属。这样一改就允许解释者在他关于相信者信念的报告中遗漏某些信念。

拜恩强调，尽管有这样的修改，但这种弱化理论仍属解释主义，因为它坚持认为，只要信念归属出现在了解释者的报告中，那么相信者就一定有此信念。

再看彭里鹏（D. Pereboom）的强硬（robust）非还原物理主义。根据他自己的判释，这是一种“关于特殊心理实在和解释的理论”。很显然，它是戴维森个例或非还原物理主义的继续和向强硬方向的发展。其动因是要解决困扰非还原物理主义的副现象论难题。根据这一理论，心理实在有特殊的因果力。所谓因果力即据以产生结果的东西。例如某人要赶一趟火车的行为，只能用某些心理状态从心理上予以解释，心理状态之所以有此作用，是因为它们是某种心理属性的例示。^⑩他不否认，心理因果力有物理的实现，

即前者为后者所承担或实现。但是它们毕竟不是同一个东西。他说：“同一性是一回事，没有同一性的实现或构成是另一回事，如果存在着这样的基本的心理因果力，它们有物理上的实现，那么任何这样的个例与微观物理实现基础的关系就不是同一关系，而是实现或构成关系。”强硬的非还原就体现在这上面。^⑪

另外，解释主义还有一种泛化的形式，即将它的基本原则推广到社会学、文化学、认识论（真理观）、科学哲学等之中，以对广泛的问题做解释主义说明。它的宗旨是，根据解释主义说明知识和实在（社会、文化、自然等）。基本观点是：实在不能与我们关于它们的认知、知识分离开，研究者、认知者的知识、价值观、世界观渗透到了全部研究过程及其结果之中，以至于浸透在所认识的实在之中，真理是通过对话而完成的协商的产物。根据这一理论，认识论必须予以重建，即必须建立在解释主义的基础之上。如此重建的新认识论的基本结论是，我们无法把我们自己与我们所知的东西区分开来。研究者与其对象如此联结在一起，以至于对我们是谁，我们怎样理解世界的回答，都成了我们怎样理解我们自己、他人以及世界的组成部分。

三、更具整体论性质的解释主义

下面再看克里格尔（U. Kriegel）的“更具整体论性质的解释主义”。他的出发点是丹尼特的解释主义。在丹尼特看来，人们之所以能对他人行为做出解释，是因为他们有意向概念如信念、愿望等组合在一起的网络。一旦人们用意向概念对人的行为做了解释，就等于对解释对象采取了意向立场，这一来，被解释的对象就被看作是意向系统。质言之，某人所具有的相信 P 之类的状态，不是因为他身上真的出现了这样的状态，而是因为解释者把这状态归属于他。同理，一意向状态有意向内容 C，就等于那状态的主体被解释为处在有 C 的意向状态之中。丹尼特说：“成为一个意向系统说的不外是，真实地、真正地相信 P，对于这系统来说，P 在最充分（最具预言性）的解释中作为一个信念出现了。”因此人有无信念是一个解释问题，而非事实问题。^⑫这种解释主义的基本倾向是自然主义，即不承认意向性是超自然现象，认为可根据更基本的物理属性来予以说明。就其相对于物理属性而言，现象性意向性

又是派生性现象。^⑨

克里格尔当然没照抄丹尼特的解释主义，而做了自己的发展。他强调：即使是非派生性意向性，只要经过了解释，就一定有其派生性，即某人被解释有某意向状态及内容，来自解释者的意向状态及内容。因为“解释某人处在某意向状态中，本身就是（解释者的——引者）一种意向行为，这行为的内容是：某人（被解释者）处在某意向状态之中。因此丹尼特的解释主义意味着：每一意向状态的意向性都来自另外某个意向状态的意向性，即是说，来自有关的解释状态（他自己的内容大概又来自第二阶解释行为的内容）。总之，丹尼特观点的结论是：所有意向性都是派生性的。”与其说这个结论是丹尼特的，不如说是他自己的创造性发挥。^⑩克里格尔认为，这个结论面对的困难是：会陷入无穷后退，即某人有意向性依赖于解释者的解释行为，这解释行为作为有意向内容的意向行为又来自于别的解释者，如此类推，以至无穷。能否跳出这个无穷后退呢？他的回答是肯定的，方法是：“假定存在着一类有特权的意向状态，例如（1）它们的意向性不来自于解释，（2）它们有作为子集的解释行为。”这思想当然是他对丹尼特的超越，也可看作是他的新解释主义的突出表现。^⑪但问题是：是否真的有这样的有特权的意向状态呢？他说：“可把现象性的有意识状态规定为这样的状态，它们具有作为必要子集的现象性的有意识解释行为。”“一旦它们这样做了，进而它们的一些行为成了解释的现象上的有意识行为，那么我们就说明现象上的有意识状态的意向性是来自于它们的解释。”他总结说：“使无意识事项具有它所具有的意向内容的东西……是，最好地采取意向立场就会导致一种解释，根据这个解释，那个事项获得了它所具有的意向内容”。^⑫可见，要跳出无穷后退，关键是看到有这样一种解释，它本身不能有现象意识。因为只要它有这样的意识，就一定有别的解释，以至无穷。这样的解释就是特权解释。他说：“这样的解释没有必要是有意识的，但它们又必须是有意识解释所组成的解释链的组成部分”。^⑬上述关于特权解释的思想可这样予以简化：“对于任何无意识事项 X 和意向内容 C 来说，X 有 C，当且仅当 X 是这样的，以至在某些条件下，理想解释者中出现了一种具有〈X 有 C〉这样内容的有意识意向状态。”

究竟什么是解释、理想解释者、理想条件？

克里格尔深知：在这里若不对它们做进一步阐释，再努力予以简化也是无济于事的。很显然，解释的本质问题是解释主义的首要问题。解释主义之所以有不同版本，是因为人们对解释的本质有不同的看法。大致有三种不同理论，（1）理论理论，即认为人们据以解释行为、完成意向归属的资源是一种心灵理论。（2）模仿理论。它认为人们的解释资源不过是一种模仿能力。对意向的归属的根据是模仿被归属者内部发生的过程。（3）合理性理论。这主要是戴维森的看法。他认为，意向归属以宽容原则为基础。根据这一原则，每个人都是有理性的，或者说，被解释、归属的目标都是理性自主体。

在三种对解释本质的说明中，克里格尔倾向于模仿理论，但他认为，要形成对解释的正确解释，还应借鉴别的方案的思想。他认为，解释包含理论机制和理性机制的共同作用。更具体地说，无意识意向状态的归属是分两步完成的，第一，向最好解释的推论可导致大量可能的归属，它们完全一致于解释项。第二，可借助宽容性原则在它们中选出一个归属，使之成为对待解释目标的最好的解释。他说：“我的解释主义版本强调的是把作为解释要素的理论化和理性化结合为一体”。^⑭

理解了解释，就不难理解理想解释者。他说：“理想解释者是这样的解释者，他能在一切条件下完善地执行意向策略。”要这样，他必须有这样一些特点，首先，他作为解释者能成为进入解释状态的主体。所谓解释状态，即是把内容归属于某物的状态。其次，理想解释者还有认识上的优势，如有完善的知识，即有关于非意向事实和所有现象性意向事实的完善的知识，有能力做出每一有效的演绎推理和有根据的非演绎推理，有能力避免做出无效的演绎推理。具备这些条件，就是理想解释者。^⑮理想解释者之所以是理想的，是因为他有最好的解释资源，如有民间心理学，有有意识的意向性和无意识的意向性。他说：“由于有意识意向性，解释者便能追踪意向性，由于有无意识意向性，解释者就能构造意向性”。^⑯理想解释者之所以能恰到好处地做出意向状态及内容的归属，是因为他有无意识事项。由于有它，它就一定有这样的倾向性、潜力，即在解释者身上产生出恰到好处的解释。这具有认识论上的必然性。他说：“对于无意识事项 X 和意向内容 C 来说，X 从认识论上说必然有 C，当且

仅当有一可能的有意识事 Y，且有一可能的理解者 N，可能的条件 K，以至（1）Y 有内容〈X 有 C〉，（2）X 在 K 条件下从 N 中引出 Y”。^③

至于理想条件的本质，他认为可这样予以定义：一条件对于主体 S 和任务 T 来说是理想的，当且仅当在有内容 C 的情况下，S 尽其所能最好地完成了 T。这就是说，理想条件是相对于被解释者 S 和他所要完成的行为任务 T 而言的，指的是被解释者最好地完成其任务所依赖的条件。^④

既然他坚持解释主义，那么他就无法回避这样的解释的不确定性问题：对于同一个做出了某行为的被解释者，一个解释者在解释时归属的是内容 C₁，而另一解释者归属的却是内容 C₂。对此，克里格尔的看法是，这类不确定性极为罕见，即使出现也无妨，因为它没有什么害处，尤其是在标准的情况下，不会出现一种以上的最好解释。例如理想解释者在理想条件下所做的最好解释只能是一种。既然是理想解释者，他就一定知道一切现象性意向事实，他就能对被解释者的意向状态中的认知的、推理的相互作用做出清楚把握。

最后我们来分析克里格尔的思想的特点。他自认为，相对丹尼特、戴维森等人的理论而言，他为解释主义阵营奉献了一种“更标准、更具整体性的解释主义”。一般而言，解释主义常与整体论连在一起，如它认为，只有有办法将内容归属于整个群体时，才能把内容归属于其中的某一个人。这是因为，人的行为容许多种一贯的解释，在这些解释中，将内容归属于一个事项的变化会为将内容归属于另一事项的变化所抵消。这样的整体论体现的是行为解释的特点，因为解释主义必然会导致整体论。不过，他坚持的整体论有这样的特点，即强调存在着无意识的意向性和意向状态，因此他的整体论包含有关于无意识内容的整体论。他说：“关于无意识意向性的解释主义必然导致关于无意识内容的整体论”。^⑤与这种解释主义连在一起的是这样的观点，即归属中的重大错误和不合理性是先天不可能的。这是因为，当几个有一致性的解释出现了时，在它们中的选择就会诉诸宽容原则。而这一原则与整体论相结合，就会阻止人们将错误的、不合理的内容归属于某事项。^⑥

其次，他的解释主义既坚持一般解释主义所坚持的关于意向性的非实在论或工具主义，同时，在无意识心理问题，他又有实在论倾向。非实

在论的基本观点是：意向性是有用的虚构。这意味着，它能帮人们对行为做出有用的解释和预言，但它并没有真实的存在地位，就像引力中心、经纬线一样。克里格尔的解释主义承认这一点，但他同时又强调：对无意识意向性的归属一方面有有用性，另一方面又有真实的存在地位，因此不是虚构，而且有用性和真实性必须结合在一起理解。^⑦他说：“促使我提出这一观点和动机的是：只有这样，才能对无意识意向性做出最好的说明。”其他的说明都行不通。主要有三种。第一，潜力论，如塞尔认为，无意识状态之所以有意向内容，是因为它潜在地包含有现象特征。第二，推理论。霍根等人认为，使无意识状态有意向内容的东西是，它从推理或功能上整合进这样的认知系统之中了，其中，有些意向状态是现象上有意识的。换言之，使它有某一特定意向内容的东西是它的特定的推理或功能作用。第三，取消论。基本观点是否认存在着无意识意向性。这些说明的问题在于：几乎抹杀了无意识意向性的存在地位。取消论自然是直接否定，其他两种理论表面上承认了，但由于没有给予适当的说明，实质上等于取消了它的存在地位。^⑧克里格尔的新解释主义强调：应回答到解释主义这样的结论上来，即某种无意识状态有意向性。这是意向性的一个特例。相信这一点的根据有两方面，一是认知科学的成果，二是民间心理学的信念。在认知科学看来，承认某种无意识状态有意向性，从解释上说是有利的，当然他承认，我们没有办法亲知无意识的意向状态，“我们相信它们的唯一理由是：相信它们存在，从第三人称观点看既有理论上的利益，又有解释上的好处”。^⑨

综上所述，新解释主义不仅是心灵哲学中的奇葩，特立独行，新论迭出，而且在解释主义阵营中也保持着自己的鲜明个性。它们既提出了许多不同于老解释主义形式的新颖的思想，对解释主义作了大刀阔斧的改革，而且也为之做了新的阐释和发展，如用“理想的归属”诠释解释主义的核心概念“解释”，并对解释主义所关注的一系列问题和所提的一系列观点做了新的论题。在此基础上，还对解释主义所面临的难题以及人们的诟病作了回应，如对人们从心理因果性、现象学性质、自知等角度提出的问题和责难做了新的解答。就问题而言，新解释主义在三个问题上与老解释主义分道扬镳了。这三个问题是：解释中

的理性与非理性问题、解释的根据问题、是否存在解释所假定的意义概念。就关注的范围言,新解释主义关注的范围扩大到了一切心理现象,因此是关于心理现象的解释主义,而丹尼特和戴维森似乎只关注意向状态。再者,丹尼特和戴维森解释主义是方法论或认识论类型的解释主义,不是形而上学的解释主义。因为他们没有对命题态度作构成性说明,而新的解释主义一般有较浓厚的形而上学色彩,如对心理现象作了构成性说明。

解释主义中,有的形态十分接近取消论,有的甚至是巧妙的取消论,但多数形态是反取消论的,而且是介于取消论和实在论之间的。取消论认为,人们对心理的常识理解可称作民间心理学(一般简称为FP)。它是根本错误的,因为根本就没有FP所相信存在的心理状态。心理状态不过是FP的理论构想。根据常见的阐释,取消论以指称理论为其前提,或是从中推论出来的。根据指称论,错误理论的术语是没有指称的,心理术语也没有指称,因此FP是错误的,是应予取消的。米勒认为,取消主义并不是依赖于指称理论的。他说:“取消主义没有必要被看作是依赖于指称理论的东西。”这也适用于包括归属论在内的解释主义。它们都没有必要诉诸指称理论,“指称的概念对于这种立场的阐述是没有必要的”。^⑤

根据米勒的诠释,取消论有概念取消论和本体论取消论之别。前者比后者更激进,它不仅否认有心理实在,而且主张抛弃心理概念,进而主张用科学概念全面替换心理概念。这种理论没有为膨胀的心理实在论留任何地盘。一些新解释主义形式赞成取消论关于心理实在所做的部分论断,如认为FP所说的心理实在是根本不存在的,应予取消。但区别在于:一是不主张取消心理概念,二

是承认FP是有用的。就此而言,它们又认为,取消论是站不住脚的。

另外,取消论像还原论、归属论、实在论等一样面临着一个共同问题,即心理状态是否是自然类型?这个问题又有两个方面:(1)心理状态是否是自然类型?(2)心理语词是否指称自然类型?两个问题的区别在于:承认心理状态是自然类型的人也会赞成说:心理语词可以指称自然类型,但在指称问题上不表态。取消论的看法是否定的,即不承认心理状态是自然类型,因此得出了取消论的结论。新解释主义认为要回答这一问题,首先要弄清“自然种类”是什么?有的认为,具备下述三个条件的事物是自然种类。(1)客观的,它们不是相对于观察者而言的。(2)是范畴,一个类型不会逐渐转化为某种别的类型,也不会从别的某型转向而来。例如水就是水,水不会是火,对于它的类型—成员问题,总有一个确定的回答。(3)内在的,类型之间的差异源于它们的构成要素的内在特征。^⑥根据这些条件,有的论证说:心理状态不是自然类型。丹尼特把主张心理状态是自然类型的观点称作“有病的实在论的堡垒”。^⑦一些新解释主义理论尽管承认心理状态不是自然类型,但又不赞成取消论由此推出的取消论结论。

新解释主义不同于取消论的特点还在于:主张FP的术语是解释主义不可缺少的,而且认为它们是有意义的。其意义可用常识心理的角色来说明。如果是这样,那么当我们对心理状态做出归属(如说某人具有某信念)时,我们就是把某角色或作用归于那个人。^⑧在归属论看来,心理语词有意义,这是最起码的事实,是没有争论的。例如在出现了某心理语词的地方,该语词就会获得它的意义。说“S相信P”意味着这样的事情,即S有条件得到承诺P的知识。^⑨

① Dennett D. “Consciousness Explained,” 转引自 Bo Dahlbom (Ed.), *Dennett and His Critics*, Oxford: Blackwell, 1993, p. 13.

② Dennett D. *The Intentional Stance*, Cambridge: The MIT Press, 1987, p. 342.

③ Audi R. “Mental Causation: Sustaining and Dynamic,” in Heil J et al (Eds.), *Mental Causation*, Oxford: Clarendon Press, 1993, p. 54.

④ Davidson D. “A Coherence Theory of Truth and Knowledge,” in Lepore E (Ed.), *Truth and Interpretation*, Oxford: Blackwell, 1986, p. 315.

⑤ Johnston M. “Objective Refigured,” in Haldane J et al (Eds.), *Reality, Representation, and Projection*, Oxford: Oxford University Press, 1993, p. 126.

- ⑥ De Caro M. “Davidson in Focus,” in De Caro M (Ed.), *Interpretation and Cause*, Dordrecht: Kluwer Academic Publishers, 1999, p. 112.
- ⑦ Brook A. “D. Davidson, D. Dennett and the Origins of the Normative Model of the Mind,” in Bailey A (Ed.), *Philosophy of Mind: The Key Thinkers*, London: Bloomsbury, 2014, p. 189.
- ⑧ ⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ Mölder B. *Mind Ascribed*, Amsterdam: John Benjamins Publishing Company, 2010, pp. 198、197、85、80 – 21、83、82、82、142、145、150.
- ⑩ Davidson D. “Indeterminism and Antirealism,” in Kulp C B (Ed.), *Realism/ Antirealism and Epistemology*, Lanham: Rowman & Littlefield Publishers, 1997, pp. 109 – 122.
- ⑪ Byrne A. “Interpretationism,” in Casati R, Tappolet C (Eds.), *European Review of Philosophy*, 3: *Response – Dependence*, Palo Alto: Center for the Study of Language and Information Publications, 1998, p. 3.
- ⑫ ⑬ Pereboom D. *Consciousness and the Prospects of Physicalism*, Oxford: Oxford University Press, 2011, pp. 127 – 128、135.
- ⑭ Dennett D C. “True Believer,” in Heath A F (Ed.), *Explanation*, Oxford: Oxford University Press, 1981, p. 72.
- ⑮ Bayne T. *Cognitive Phenomenology*, Oxford: Oxford University Press, 2011, p. 82.
- ⑯ ⑰ ⑱ ⑲ ⑳ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞ ㉟ Kriegel U. “The Basis of Unconscious Content,” in Bayne T et al (Eds.), *Cognitive Phenomenology*, Oxford: Oxford University Press, 2011, pp. 83、83、83 – 84、84、85、86、87、87、87、88、88、89、89、90.
- ㊱ Ellis B. *Scientific Essentialism*, Cambridge: Cambridge University Press, 2001, pp. 19 – 21.
- ㊲ Dennett D. “Get Real,” *Philosophical Topics*, 1995, vol. 164, no. 28), pp. 72 – 74.

(责任编辑: 颜 冲)