

免于误认的错误与自我知识的 行动主义说明*

安秀媛 (AN Xiuyuan)**

摘要：我们如何能对有关我们自身的知识拥有确定无疑的认知辩护一直是心灵哲学中的重要话题。免于误认的错误 (immunity to error through misidentification, 简称 IEM) 论题在解释自我知识时的基本观点是, 对于有自我意识的主体“我”而言, 某些自我归属的有关第一人称内容的心理状态是不会发生识别错误的。支持该论断的说明方式之一是行动主义 (agentialism)。该说明认为, 我们之所以能拥有关于我们自身的信念和意图, 正是因为我们能对它们行使行为能力 (agency)。然而, 行动主义对主体理性的行为能力的要求会受到来自情绪等非理性因素干扰、主体归属不一致以及不同种心灵观导致不同的自我指称的挑战。本文将考察行动主义说明, 并以两种回应策略来支持一种较行动主义更弱, 但适用范围更广的理性主义说明。本文将论证, 行动主义说明对主体意识的要求过强, 我们应当采取一种更为温和的理性主义, 以便既能满足自我知识的认知可靠性, 又可以解释第一人称权威。

关键词：免于误认的错误 (IEM); 自我知识; 行动主义 (agentialism); 理性主义; 自我意识

* 特别感谢张小星教授对本文早期版本的指导和鼓励, 感谢他为本文的写作、论证思路和内容提供非常有帮助和建设性的评论。非常感谢尹洁教授对本文观点的启发和充满洞察力的建议, 也非常感谢董自立博士和刘可萱对本文选题的探讨、推荐相关参考文献以及热情的反馈。

** 安秀媛, 复旦大学哲学学院硕士研究生, 电子邮箱: 22210160045@m.fudan.edu.cn。

Immunity to Error through Misidentification and the Agential Approach to Self-Knowledge

Abstract: The question of how we can be certain that we have an epistemic justification for our self-knowledge has been among a central debate in philosophy of mind. The basic idea of the immunity to error through misidentification thesis (IEM) in explaining self-knowledge is that certain self-ascriptions of mental states concerning first-person content are not prone to error for the self-conscious “I” - thoughts. One of the accounts that supports that claim is agentialism. According to this view, we can have beliefs and intentions about ourselves because we exercise agency over them. However, irrational factors such as emotional interference, inconsistent subject affiliation, and different views on the mind leading to different self-references can call this claim to the subject’s rational agency into question. This paper examines the agentialism, identifies its flaws, and presents two alternative strategies to support a rationalist account of self-knowledge—which is generally weaker than agentialism. It argues that the agential approach to self-knowledge is too demanding on the subject’s consciousness thus a more moderate approach is required to satisfy both first-person authority and the epistemic reliability of self-knowledge.

Key words: Immunity to error through misidentification (IEM); Self-knowledge; Agentialism; Rationalism; Self-consciousness

— 免于误认的错误与自我知识

当代心灵哲学的认识论问题的重要研究方面之一是为心灵的特殊性和认知的不可错性辩护。当我们把自己以一种主体的方式理解并指称时——例如当我说出或想到“我牙疼”“我感到快乐”——我不会把其他人误认为我自己。像这样有关我们自身的第一人称思想是不会出现识别错误的。这种情况被称为免于第一人称代词误认的错误 (immunity to error through misidentification relative to the first - person pronoun), 简称免于误认的错误^① (immunity to

^① 亦可翻译为“对第一人称代词的错误识别是免疫的”和“免于错误识别的错误”，为了简洁起见，我选择了文中的译法，并在适当时使用其缩写 IEM。

error through misidentification) 或 IEM。^①

当代哲学家对自我进一步做出了更细致的划分：他们将驱使我们有意地进行反思活动的这个有意识的主体，它所具有的将它自己（即“我”）从其他东西中识别出来的能力称作自我意识（self-awareness/self-consciousness），并把有关这个主体的一切心智活动的知识称作自我知识（self-knowledge）。前者的例子有，“我”进行自我定位、将感官的感知进行有效的归属，拥有信念、态度、能够做出判断的能力等；后者的例子有，“我”对“我”所意识到我的思想、经验、欲求、信念的知识。自我意识与自我知识二者紧密相关，一个人关于自我知识与自我意识的观点也对其理解心灵、知识与实在有着基础而深远的影响。不难看出，这使得在解释心理活动时，特别需要我们的第一人称以及主体起作用。

在对 IEM 的诸多说明中，理性主义（rationalism）认为我们的拥有自我知识能力来自我们对理性的运用：我们能够使我们的态度符合我们的理由，并能通过证言（avowal）确认自己的态度。与理性主义相近的行动主义（agentialism）则认为，不仅是我们的理性能力保证了我们的知识，我们还应当对其持有一种主动的态度，甚至连我们拥有的自我知识的能力都是从我们的理性行为能力（rational agency）中派生出来的。^②这两种理论都认为，正是我们作为理性思考者的地位保证了我们的自我指称和自我知识的可靠。其中，行动主义尤其认为，这种保证来自理性思考者在理性慎思中对其行为能力的运用。需要说明的是，哲学家们一度把对于理性和理性行为能力的运用统称为理性主义。而理性主义作为哲学史上最具有影响力的观点之一，人们对它的内容和蕴含是不断修正的。行动主义正是产生于这样一种契机，它也可算作一种广义的理性主义。我们在这里只需明确，当代行动主义者通常也是理性主义者，而非反之。为了表达清晰，下文所使用理性主义是指那些专注于理性（rationality）并独立于行为能力（agency）的理论。这就是说，本文在解释自我知识时所使用的理性概念并不与行为能力相捆绑。^③这也是使得我们能从行动主义转到更为一般的理性主义的理论前提。

① 参见 Anscombe (1981), Shoemaker (1968: 557), Wittgenstein (1958)。

② Gertler (2018: 91). 参见 Bilgrami (2012), Burge (1996), Moran (2001), Shoemaker (1994)。

③ Agency 指的是个体基于其自由意志有意图地进行活动的的能力，还可翻译为“行动能力”“行为能力”或“能动性”。为了避免与中文语境下的其他用法相混淆，本文选择了“行为能力”这个翻译。

然而，无论是身患精神疾病者，还是处于激情和情感蒙蔽了理性判断状态下的常人，人们的理性控制作用经常会失效——那我们是如何获得对我们的自我知识的辩护呢？本文旨在为自我知识寻找坚实的理论说明，并指出现行的行动主义要求过强，使得我们可以无法得到一个整体而具普遍性的解释。因此第三节以体现着更为一般化理性主义的反身规则和笛卡尔原则两个策略，回应行动主义面临的困境。最后给出我对理性主义的自我知识可靠性的辩护。

二 行动主义说明及其困境

被普遍接受的行动主义观点是，行为能力一般包括影响某对象或者现象，而知识一般包括被一些对象或现象所影响的东西。在此基础之上，行动主义者还接受这样的观点，即，对感觉和其他被动心理状态的自我知识是有关观察（或以其他方式跟踪）这些状态的事情。^①但同时他们认为，一个主体不能仅仅通过观察他们的认知生活，就可将他们自己的信念和意图理解为“规范承诺的主动状态”。^②出于这个原因，行动主义者认为关于信念和意图的自我知识与关于被动心理状态的自我知识有很大不同。^③正是在这个关于相信和意图某事的理由上，行动主义与理性主义产生分歧，因为后者认为这两者对于认知权威而言没有区别。

那么，首先让我们用行动主义的说明来解释精神分裂症患者的口头报告。

【例 1】有一位女性患者（代称 F. B.），在右脑中风后，出现了偏侧感觉缺失（hemianesthesia）和身体妄想症（somatoparaphrenia）（一种精神分裂症）。她对自己的左边身体没有感觉，她在被要求触摸她的左手时，也只是在身体右侧去寻找，但无法够到左手。此外，F. B. 还认为她的左手属于她的侄女。于是实验人员对 F. B. 进行了新的测试，内容有三项：实验员用食指短暂地触摸 F. B. 的手背，并告知她，他/她将会触摸（a）她的右手，（b）她的左手，（c）她侄女的手。在后一种情况

① Gertler (2022: 253 - 54).

② Korsgaard (2009: 39).

③ Gertler (2022: 254).

下 (c)，实验员触摸了 F. B. 的左手。F. B. 被蒙住眼睛，被要求在感觉到触摸时说“是” (yes)，在没有感觉到任何触摸时说“否” (no)。测试结果是，(a) 的每一次回答都正确，(b) 是都不正确 (她都没有感到触摸)，而 (c) 则是大多数情况都答对了。在为期两天的测试中，第二天的正确率较第一天相比，有显著提高趋势，最终正确率为 70%—100% 不等。^①

正如我们可以从 F. B. 的例子中注意到的，即使是精神分裂症患者也拥有一定程度的理性能力，从而能够对其感知进行归属。并且他们的感知能力是受到更高一级的主体感知所控制的，在受到损伤后可以借由语言上的指导得到一定程度的恢复。后一点似乎意味着，能对思想进行控制、有行为能力，对于主动认知而言也是非常重要的部分——该理论蕴涵也可以为自我知识的行动主义说明所接受。

有鉴于此，行动主义者科利瓦 (Annalisa Coliva) (2012) 给出了一种关于承诺 (commitment) 的自我归属的观点，强调了理性持有的主动态度 (如欲望)：假设一个思想者是理性的并且有相关的概念，那么他们对承诺的自我归属将是真实的，因为他们会创造归属的态度。^② 但就自我知识而言，行动主义说明基于对人类认知的过于理想化描述，不可避免地会面临一些问题。

第一，行动主义说明重视信念和意图的主动控制，但这个主动性本身可能是不被保证的。一种情况是，我们塑造理性和态度的可能是非理性因素，这会对能说明自我知识的理性行为能力的认知基础造成挑战。一个更直观的例子是，一位父亲要决定是否借钱给女儿开餐馆，虽然在理智层面他无法预测这个本金是否能得到归还，但在情感层面上他拥有希望女儿快乐幸福的期盼——正是这个期盼使得他愿意去相信女儿能够经营得当，获得事业上的成功。所以他的行动判断在认知特权上并不比其他一般信念更具有优先性。更重要的是，由于信念的透明性，是他的信念塑造了他对其理由的感觉，而非他对其理性行为能力的控制形成了他的信念。

回到 IEM 所要归属的自我，它同时具备心灵和物理两种属性，但无论持有二元论还是观念论的自我，都需要接受这个自我是寓于物质性身体之中

① Bottini et al. (2002: 249–51).

② Coliva (2012: 235–36).

的，并且是通过一系列物理属性来呈现的。换言之，行为能力本身并未保证自我归属中所指自我一定是纯粹的精神实体，因而也仍不可避免地存在身体性误认的可能。如果行动主义退而求之，选择心理自我归属不会发生误认，就会使得它与理性主义对信念的观点差别不大了。这是因为，那些不依据理性控制就可产生的心理活动会被行动主义视为一种被动的感觉——因为主体仅仅是经历到一种行动并不能使其成为主体的行动——从而也更不会被行动主义视为自己理论所独有的部分。我们将在第三节继续讨论自我指称和自我归属的同一性问题。

第二，既然我们并非完美的理性认知者，有犯错的可能，那么我们对于运用理性思考的行为能力也并非一直无可指摘。如果我们接受理性主义的弱观点成立，那可以接受的观点是，思想的特定现象特征决定了有关这个现象特征的内省状态，二者的构成性关系保证了理性认知的可靠性；但行动主义对理性所持有的主动态度却无法仅仅依靠认知可靠就被建立。更何况，理性认知允许我们以第三人称的视角对我们的周遭世界进行审视思考，认知并非在所有时刻都是以主体为中心发出的。那么强调主体对其自身的行为能力的控制也无法保证作为认知结果的自我知识的特殊性。

第三，行动权威 (agential authority) 是全有或全无的事情，而认知特权 (epistemic privilege) 是程度的事情。一方面，行动权威的独特之处在于它结合了一个人自己的理由—责任态度 (reasons—responsive attitude) 与行使其行为能力；而认知权威可以延伸到主体经历的任何现象上。^① 这使得理性主义在认知权威上可以满足自我意识的主体性要求，而行动权威就不行。另一方面，接受行动主义的强观点还会带来危害其自我知识的新的理论蕴涵：除非我们知道促使我们行动的态度，不然我们无法对我们的行动负责。一旦我们对自己相信某事或判断某行为的状态是不确定的，那行动主义就无法在概念上拥有对行动的承诺，因而也就无法解释该主体的行为理由了。

我们可以对上述三点做一个小结。对自我知识的特殊性的解释依赖着对具有 IEM 现象的自我意识的说明，这个理由同时也解释着心理和身体活动。因此，即使可以暂行悬置对自我的概念的形而上学设定（是二元论、观念论抑或是物理主义等），也需要保证自我意识能够满足指称时的同一性，即，使其真正成为主体的有意识的思想；不然则不会对 IEM 形成挑战。本文给出

^① Gertler (2022: 256).

解决问题的出路是：如果能对自我意识给出一个有规范性和具限定作用的条件，那么在这个范围内便可使主体获得可靠的知识，即自我知识。

三 理性主义：对行动主义困境的回应

上文介绍了行动主义在解释自我知识时面临的诸多困难，但这并不意味着我们不得不接受一种经验性的可错的自我。我们指出，产生这些困境的原因，并非因为由于理性和主体性本身无法应对复杂的自我意识，而是对行为能力提出的过高要求所致。本节用与行动主义共享部分理论基础但不要求行为能力的一般理性主义为自我知识的可靠性辩护。理性主义对行动主义困境的回应是，以反身规则策略保证自我指称时自我的同一；并以笛卡尔原则策略保证这个思想是有意识的，即处在理性认知范畴之下。

（一）反身规则策略

关于自我知识“我”的指称问题的传统解释之一认为，有关第一人称的指称是固定的（fixed），它被反身规则（reflexive rule）所决定，即，任何第一人称思想的个例（token）都是关于这个思想的思想者的。^① 所以它又可称作个例 - 反身规则（token - reflexive rule），意思是说，“我”的任何个例都指称（指向）其发出者。^② 我们意图说明，行动主义由于其对行为能力的强调，导致了自我归属困难；但借由指称的反身规则，一般理性主义就足够回应 IEM 问题有关自我的困境。

一般认为，自我意识（self - consciousness）指的是心灵的一种能力，比如某主体有自我意识；而有自我意识的（self - conscious）是有关心理状态的一种特殊的属性。因此，反身规则可以使得拥有自我意识的属性从一般思想命题中区分出来。以下面【例2】俄狄浦斯（Oedipus）的一组思想为例。^③

【2.1】解开斯芬克斯之谜的人杀死了拉伊俄斯（Laius）。

【2.2】我杀死了拉伊俄斯！

在【2.1】和【2.2】中，只有【2.2】是有自我意识的思想。这意味

① Palmira (forthcoming). 参见 Howell (2001: 127), Perry (1997: 595 - 596)。

② Campbell (2012: 6)。

③ Echeverri (2020: 473)。

着,当俄狄浦斯想到(或说出)“我杀死了拉伊俄斯!”的时候,他意识到自己先前的行为所带来的后果,并开始需要承担道德责任。IEM 问题在此处的体现是,一方面,他不可能在合理地相信【2.2】的同时怀疑“我”所指称的究竟是不是他自己;另一方面,俄狄浦斯却可以分别相信【2.1】和【2.3】“我解开了斯芬克斯之谜”,并怀疑【2.4】“‘解开了斯芬克斯之谜的人’是否指称的是我呢?”——如果他忽视“我 = 解开斯芬克斯之谜的人”这一事实的话。

根据反身规则,只要拥有受反身规则的概念类型决定的第一人称指称,就可以让拥有该思想的思想者以拥有自我意识的方式被解释。然而,通过例【2.1】到【2.4】我们很容易发现,虽然“我 = 解开斯芬克斯之谜的人”同样是关于俄狄浦斯这一主体,但是它们对自我指称确有着不同的意义:对于前者“我”,第一人称指称的解释被反身规则所决定,但对于后者“解开斯芬克斯之谜的人”却不具备任何解释效力。尽管反身规则本身无法为自我意识提供充分条件,但我们也不必放弃用这个规则去解释第一人称的使用情况,特别是来解释自我知识。这是因为,反身规则为我们强调了自我意识在意识的指称活动中所扮演的重要作用,它可以帮助我们分析思想者的认知问题,以及它是在何种意义上对自我知识的可靠性有效。

埃切维里(Santiago Echeverri)和帕尔米拉(Michele Palmira)将形如【2.1】和【2.4】中“解开斯芬克斯之谜的人”称作 *de facto* 反身 (*de facto reflexive*),将【2.2】中的“我”称作 *de jure* 反身 (*de jure reflexive*)。^① 埃切维里注意到,有自我意识的 (*self-conscious*) 思想包括了 *de jure* 反身的概念类型,并且他认为只有 *de jure* 反身使得“我”从其他概念类型的指称中独特出来:所有“我”的个例都使得指称有了保证 (Guarantee)。这就是说,如果一个思想有保证,那拥有这个思想的思想者就不能在拥有这个思想的第一人称概念时,对这个个例的指称产生疑问。^② 这与自我知识的 IEM 现象的描述十分相似,因此很多人在说明指称问题上把应用了反身规则的保证与 IEM 合并为同样的问题。^③

① Echeverri (2020: 478).

② Echeverri (2020: 476 - 477).

③ Cappelen and Dever (2013: 133 - 135), Rovane (1987: 147 - 167). 此处还有舒梅克、普莱尔、克里斯托菲杜 (Andrea Christofidou) 与科利瓦关于类记忆及其对应的免疫的自我知识的争论,但其不是本文所要讨论的目标。参见 Shoemaker (1970: 283), Pryor (1999: 289), Coliva (2003: 420), Christofidou (1995: 223 - 234)。

正如我们在第二节的精神病理学案例中看到的，包括内省和内感觉在内的所有有关自我知识的信息来源都受到了程度不一的反驳或被举出了反例。这些反例或者从记忆，或者从本体感（*proprioception*）、身体的其他知觉或行动状态中派生出了异常的因果链，使得本应独属于我自己的意识经验拥有了其他人的信息输入，进而使得作为自我意识特殊性的 IEM 受到挑战。我希望指出的是，这并不意味着这些异常情况对所有 IEM 问题都构成挑战：在声称某个案例成功挑战了 IEM 之前，必须先明确它挑战的是哪种类型的 IEM 以及它所指称的是自我的哪个方面。

例如舒梅克（Sydney Shoemaker）提出了间接推测免疫（*circumstantial IEM*）和正常情况下的绝对免疫（*absolute IEM*）。前者的判断依据其所处环境而改变，比如在光线不沿直线传播的非欧空间里，依赖视觉的指称判断就是可错的。^① 又如普莱尔（James Pryor）区分了两种不同的误认，以及相对应的免疫：*de re* 误认（*de re misidentification*）与 *wh* - 误认（*which - object misidentification*，或称 *wh - misidentification*）。^② *de re* 误认要求存在 *de re* 信念，而 *wh* - 误认则包括对存在命题（信念）的识别。但问题是，判断误认类型有时可以很困难，因为主体的自我识别经常是模糊的。

从【例 1】里我们可以注意到，由于 F. B. 可以同时认为“我侄女不是我，且与我身体不同”，并且她声称感受到触摸的其实是她的侄女，所以在她看来，应当进行识别的其实是“她的侄女”而不是她。即，第一人称视角的现象经验并不一定都被表现为以第一人称方式经历的。^③ 然而，不存在是她还是她侄女的中间状态，如果对作为她侄女的程度有所怀疑，这时候也会感受不到任何触摸（那就等于是她自己了）。F. B. 的例子说明，身体意识来源于身体更高层次的基本信念的表征，如所有权（*ownership*）。^④

如此看来，经验归属与自我所有并不一定一致，这意味着我们自我指称的“我”究竟是指称物理世界中的身体，还是指使得感官经验得以可能的有意识自我（先验自我），不能仅仅用“先验自我具身于物质身体上”^⑤ 来简单澄清。以 F. B. 为代表的身体指涉困境表明，IEM 若要成立，所需满足的

① Shoemaker (1968: 562 - 563).

② Pryor (1999).

③ 参见 Liang (2015: 8)。

④ Bottini et al. (2002: 252).

⑤ 持有类似观点的人如埃文斯（Gareth Evans）和科利瓦。参见 Evans (1982: § 7)。

不只是概念的真理 (conceptual truth), 还要能满足一些其他的个例要求。

将 IEM 现象与埃切维里和帕尔米拉的反身规则相结合, 我们便可发现具备认知可靠性的自我知识与指称自我 (无论它是身体的抑或是先验的) 的关联: 自我意识的 IEM 现象的实质是要说明 *de jure* 反身所意指的“我”究竟希望获得一个怎样的自我同一概念。不同种的 IEM 与保证研究的对象相同, 均既可以是心理归属, 也可以包括身体经验的归属, 且均是通过讨论专名、索引词、指示代词等方式来呈现, 并且所要回应的问题同样是对自我的指称何以具有特殊性。^① 但同时需要说明的是, 无论是反身规则还是作为其说明理由的保证, 在提出之初都并不是为了直接论证 IEM 的不同版本哪种更为有效, 或者更有哲学上的认知价值的; 不如说, 它是对有或强或弱的立场的 IEM 的内部成立理由的一种说明方式。但无论怎样, 区分 IEM 的免疫对象是什么可以帮助澄清认知者在自我指称时同一性问题, 这将有助于自我知识辩护的规范性维度。

(二) 笛卡尔原则策略

很多人都同意, 有意识的思想包括现象感、主体感、本体感等。就主体感而言, 它体现为当我感受到我的手在摇摆的同时, 我意识到是我而不是其他人的手在摇摆。换言之, 如果我接受“我有一个思想 (thought)”, 对这个思想就不能说不是我的。^② 这个理性主义的意识观念的倡导者最初可以追溯到笛卡尔的我思 (*cogito*), 他称作“思想” (thoughts) 的东西在今天被称作“有意识的思想” (conscious thoughts)。一些人将这个观念总结为:

笛卡尔原则 (Cartesian principle): 如果一个想法是有意识的 (conscious), 那么它就是主体的 (subjective) (这是从“它在我看来是‘从内部’是我的”的意义上说的)。^③

通过笛卡尔原则, 人们想要说明主体对于其自我知识和主体性有着不可错的 (infallible) 认知安全 (epistemic security)。尽管仍有一部分学者认为笛卡尔原则和 IEM 仍然在处理认知安全的范围上存在差别, 因为 IEM 关注

① 参见 Echeverri (2020: 478 - 479) 以及其他有些许差异的版本 Coliva (2003; 2017: 233 - 252)。

② 参见“第二沉思”Descartes (1985: 19), 转引自 Billon (2013: 293)。

③ Billon (2013)。

的是自我知识判断中（被识别为）自身的对象的安全，而笛卡尔原则处理的是更为普遍的判断安全。^①但我认为这对于我们的分析不会造成影响，因为当一个主体对“我”做出免于误认的判断时，并不要求他们对自己有着特定的形而上学构想——无论是将“我”构想为一般人类身体、笛卡尔式的自我还是休谟式的知觉束都是可行的——只要他们能始终如一地按此方式理解他们。

然而，笛卡尔原则在两方面可能遭遇到一些反驳。一是当我们使用“意识”一词时是在什么意义上讲的；二是在心灵的多重能力中，哪种能够作为判断有意识的标准。如果我们获知我们的有意识状态多是通过内省而获得的判断，这样就使得意识是基于内省所得来的自我归属的某些状态的特征——那其实我们并没有说出什么比原表达更多的东西。但是这并不意味着我们需要放弃笛卡尔原则，更并不意味着能进行反思的理性能力对于主体性和有意识而言是不重要的。在我看来，坚持笛卡尔原则和接受自我知识的可靠性是可以相融的：如果我们先厘清受到 IEM 原则控制的自我意识究竟是由哪些部分组成的，以及在这些部分中哪个是对笛卡尔原则更为有效的，那我们就并不需要直接放弃用笛卡尔原则解释意识和自我知识的免疫特性。

关于反驳一，我们要去定义什么是“有意识”。满足笛卡尔原则的“有意识”是指信息加工的认知能力的意识，还是指作为去感知的现象意识呢？认为思想植入是 IEM 的反例的传统批评者一般会假定被植入的思想是现象上有意识的，是为思想者所及的，主体拥有对这些思想的行为能力，所以他们便认为患者的意识的主体性受到了攻击。然而，精神分裂症患者的例子或许可以为我们提供一种新的对于心灵维度的理解，既保持了对患者的口头报告的严肃对待，也不与笛卡尔原则矛盾。例如毕龙（Alexandre Billon）认为心灵（有意识）可以分为现象心灵（phenomenal mind）和空间心灵（spatial mind），后者也被称为心灵的功能维度。^②现象心灵包括现象意识、主体性（即所有感），空间心灵包括对反思意识（reflexive awareness）以及对反思意识的即时通达等。根据第二种功能解释，心理状态和意识本质上就是认知通路问题，所有感就是内部定位问题。尽管这两个维度通常是联系起来共同存

① 参见 Howell (2007: 596)。

② 毕龙的主张与查默斯（David Chalmers, 1996: 16）的主张相似，即心灵的双重面向，既指心理状态，也指现象状态或过程。

在，但前者却不可还原为后者。^①在他看来，使得我们的心灵特殊的東西恰恰在于其现象性部分，是令我们“有主体性的”生活的特征，而不仅仅是觉知到某思想（包括被植入的思想）的“在我们之内”。

我认为毕龙的解释是合理的。通过经验归属，我们可知在不发生误认的时候，一个主体会有相应的现象体验，所以我们可以暂且同意有意识的自我就同时要求了主体性，并且主体性是被现象性所蕴含的。思想植入的【例1】F. B. 有很好的反思意识，能够知道“‘我的侄女’拥有了这些感觉，而我不同于‘我的侄女’”，所以她否认（disown）拥有者的现象意识，因而缺乏主体性。换言之，尽管拥有反思意识，但在现象上仍是无意识的，所以也不是完整意义上的“主体”，不构成有自我意识的 IEM 的反例。

关于反驳二，或许我们可以将主体经验的所有（ownership）从行为能力（agency）中区分出来，并将主体拥有这个被归属的经验的思想者的方式区分为作为发起者（author）和作为接受者（recipient）。作为思想的发起者意味着生产（produce）了这个思想，作为思想的接受者意味着能考虑或者以某种方式直接意识到它。当一个人直接意识到某思想、有能力思考它并且内省到它的内容时，他们可以合理地相信自己是某个个例思想的接受者，但无须认为自己生产了这个思想。例如，当我想“我在想金星”的时候，我是这个思想的发出者同时也是接受者；而作为思想的接受者，却并不意味着这个主体同时一定也是该思想的发出者。又如：

【例3】一个精神分裂症患者可能会说：“有一些想法跑进了我的脑子——‘杀死上帝’。这些想法萦绕在我的脑海中，就好像是我的脑子在工作一样，但其实不是的，它们是一个叫克里斯（Chris）的小伙子的想法。”^②

在这个例子中，精神分裂症患者否认是其思想的发出者，也否认对其的所有，但确乎经历着这个思想。这种不对称的现象我们可以总结为发生在思想的行动者（agent/agency，或发出者 author）与所有（ownership）之间。

如果接受这样的观点：我们应当将意识与主体性联系在一起，使得笛卡

① Billon (2013: 310 - 311).

② 转引自 Campbell (1999: 609), Frith (1992: 66)。

尔原则是一个先验的真理 (*a priori truth*), 而理性主义和行动主义所关注的是概念的真理——如此则会带来理论的自相矛盾。因为精神分裂症患者似乎也还是拥有理性能力, 可以做出一般理性人所能做出的其他判断, 那他们的口头报告就与笛卡尔原则相排斥了。对此, 我们挽救笛卡尔原则的方法是给出一个没有被考虑到的预设问题: 作为拥有一个思想的主体意味着什么, 是作为拥有它的行动者还是接受者? 我的回答是, 笛卡尔原则所支持的主体性要求了该主体是其思想的所有者。例如思想植入就可以算是有意识, 是一种自我归属, 但并不是思想的自我所有——因为患者并不承认他们拥有了这些思想。这再次说明了, 思想的自我归属 (*self - ascription*) 不等同于思想的自我所有 (*self - ownership*), 而只有后者才算是自我意识, 是对第一人称指称的充分表达。

既然如此, 那用反身规则来辩护有意识心灵的可靠认知情况就明朗了: 思想植入包括对思想—行为能力 (*thought - agency*) 的误认, 但并不包括对思想—所有 (*thought - ownership*), 也就是接受者的误认, 因此对免疫原则不构成挑战。这意味着患者的思想的主体感依然是他们自己的, 而他们也会承认这一点。^①

然而, 这个回应可能再次面临一个质疑: 某个思想即使应用了笛卡尔原则, 依然可能由于不包括任何对自我同一性的指称, 从而使得它无法成为被自我所有的思想。

但正如上文第三(一)节所做的有关意识的反身规则的说明所示, 一方面, 当关注自我意识作为拥有现象体验的部分时, 自我并不同时要求也是该思想的行动者, 也并不关心这个思想是如何发出或被接受的。理由其一是, 作为现象体验的自我意识是身体的感官感知, 这里对应的同一性是所有者与发出者之间的同一; 而当我们谈论自我的同一性时, 我们所谈的指称是一种形而上学设想, 它所对应的应当是二元论、观念论或物理主义等对身心关系的理解——两个语境下使用的同一性概念属于不同的范畴。其二, 我们完全可以设想出一种并不需要考虑认知者的同一性的情境。在该情境中自我指称的对象是心理谓词, 它以一种认知能力来呈现。但这并非一种感官知觉, 并非把自我当作一种对象来呈现, 而是以自我当作主动去呈现对象的方式去呈现它遇到的其他知觉。^② 这时候我们不把自己想象成要被识别的对象, 而是

① Billon (2013: 295 - 296).

② Shoemaker (1968: 557).

当作进行识别判断的主体，从而还能顺利避免出现行动主义者科利瓦所担忧的，在 IEM 的指称问题里出现“心理状态不被人（主体）所有”与“‘我’不是一个指称性的表达违背了我们的语言学直觉”的情况。^① 更何况，作为一个思想的接受者与发起者并不是互相排除的关系，在对自我持有任何一种形而上学的构想下，拥有自我意识的人可以同时既是他们思想的发起者也是他们思想的接受者。

另一方面，无论是把免疫原则解释为一种先验真理^②，还是把基于内省的心理归属作为逻辑上的 (*logically*) IEM，把部分免疫的身体的自我归属称作事实上的 (*de facto*) IEM^③，都可为第一人称意识与指称的特权根据做出解释。前一类说法承认非物质自我的指称在识别中起着关键作用，它作为心灵构成性中的一环，解释指称的能力的产生根据，即二元论和观念论的立场；后一类则把物质身体作为认识现象中不可或缺的重要一环，随之而来的免疫不单是一种认知现象，而是做出认知判断本身的基础所在。不难发现，理性主义的自我知识理论本身无须承担辩护笛卡尔式二元论（或任何一种形而上学）心灵观的理论负担，只要满足对于一个思想的主体性掌握，就与免疫判断是可以兼容的。

综合上述理性主义策略的几点优势，我们可以对行动主义困境的回应做一个简短的小结。对于主体而言，反身规则的认知意义在于它使得认知主体能够从其认知属性中将所有感从行为能力中区分出来。从而使得判断的基础既不依赖于内省或其他感知觉状态，也不依赖于时刻需要保持的完全理性或行使行为能力的状态，而是基于我们与被内省到的这个思想的关系——后者正是在任何情况下都为我们所持有的。

① 科利瓦将这种现象称为说话者指称与语义指称的分裂 (*the split between speaker's and semantic reference*)。该现象是说，即使不知道哪个人是一个单称表达里的语义指称对象 (SR)，或者是后者被识别为错误的对象，依然允许一个人使用那个表达式去指称 (SPR) 某个除它的语义指称对象之外的某人。举个例子，你可以用“Coliva”去正确地指称这个 SPR - SR 分裂观点的提出者安娜丽莎·科利瓦，但实际上，由于你并不知道 Coliva 是哪个人，就把那个人误认为是她，并用这个名字来指代那个人。参见 Coliva (2003: 425 - 426)。

② 埃文斯认为“我”指称了一个先验的自我意识，是原则上不为感性经验所及的实体，由于它是无涉识别 (*identification-free*) 的，所以不会出现误认。但这个自我是具身于时空当中的，故而必须导向经验中的对象——一种现实的元素。参见 Evans (1982: 256)，Shoemaker (1968: 556)，Cassam (1997: § 2.7)，Alweiss (2018: 34)。

③ Coliva (2012b: 25)。

四 结论

通过反驳行动主义的一些强观点，本文试图说明，即使 IEM 是一个程度概念，在不同的主体的不同意识强度和不同环境下具有不同的特征和类型，我们在证明有其特征的知识自我知识的可靠性时也应给出更为普遍的解释。与此同时，以上与第一人称相关的自我意识、自我指称和自我知识，都是不可分离开独自谈论的。

我们可以认为，*de jure* 反身的自我意识会在认知上保证发生一个思想的现象特征的反思状态，是由使用该思想的这个接受者所拥有——理性主体和其自我意识要指称的自我是同一的。即使该主体不是思想的行动者，他们仍会以被动参与的形式经历到它并获得相应的现象信息，从而能够对主体性进行确认与报告。然而，行动主义所要求的理性运用的行为能力的控制过强了。自我知识在有自我意识的心理归属与身体归属上不会发生误认，是通过理性认知对现象经验的接受所保证的；但对于理性认知所带来的信念的掌控，并没有得到自我意识的保证。

尽管如此，理性主义对于自我知识的 IEM 现象的说明也可能会再次面临质疑，人们可能认为这样的解释听上去太过随意，以至于对结论的平淡感到不耐烦。我承认，即使理性主义相较于行动主义而言立场已然更弱，它仍会面对无法获得自我知识的情况。比如说自我蒙蔽（self-blind）是否可能、自我欺骗（self-deception）为何会产生、对意识的主体性的诉求是否真的是必要的，等等。这就需要我们另文去讨论理性主义对自我知识说明的合法性，而非本文解决以认知主体的行为能力为核心的行动主义解释自我知识的认知可靠性的意图所能及的了。

目前我们的结论是：在经过了反身规则和笛卡尔原则两个策略的说明后，一般理性主义对有关自我知识的第一人称的免疫原则仍然是有效的；而行动主义在宣告我们有对自身的行为能力的控制时所暗示的自我就是与物质身体的同一，并不能使得自我意识自动获得指称时的特殊性。不过，归功于他们的尝试和不断地探寻，我们得以在葆有对心灵的特殊性的好奇中，暂时离开对理性与行为能力的盲目狂热、悬置对确定性的一厢情愿，在概念与经验中感受自我意识的不同魅力。

参考文献

- Anscombe, G. E. M. , 1981, “The First Person”, *Metaphysics and the Philosophy of Mind: Collected Philosophical Papers*, Oxford: Basil Blackwell.
- Bilgrami, A. , 2012, *Self-Knowledge and Resentment*, Cambridge, MA: Harvard University Press.
- Billon, A. , 2013, “Does Consciousness Entail Subjectivity? The Puzzle of Thought Insertion”, *Philosophical Psychology*, 26 (2): 291–314.
- Bottini, G. , E. Bisiach, R. Sterzi, and G. Vallar, 2002, “Feeling Touches in Someone Else’s Hand”, *Neuro Report*, 13 (11): 249–252.
- Burge, T. , 1996, “Our Entitlement to Self-Knowledge”, *Proceedings of the Aristotelian Society*, 96 (1): 91–116.
- Campbell, J. , 1999, “Schizophrenia, the Space of Reasons, and Thinking as a Motor Process”, *Monist*, 82 (4): 609–625.
- , 2012, “On the Thesis That ‘I’ Is Not a Referring Term”, *Immunity to Error through Misidentification: New Essays*, edited by F. Recanati and S. Prosser, New York: Cambridge University Press.
- Cappelen, H. , and J. Dever, 2013, *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*, New York: Oxford University Press.
- Cassam, Q. , 1997, *Self and World*, Oxford: Oxford University Press.
- Chalmers, D. J. , 1996, *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.
- Christofidou, A. , 1995, “First Person: The Demand for Identification – Free Self – Reference”, *The Journal of Philosophy*, 92 (4): 223–234.
- Coliva, A. , 2003, “The First Person: Error through Misidentification, the Split Between Speaker’s and Semantic Reference, and the Real Guarantee”, *The Journal of Philosophy*, 100 (8): 416–431.
- , 2012a, “One Variety of Self-Knowledge: Constitutivism as Constructivism”, *The Self and Self-Knowledge*, edited by A. Coliva, Oxford: Oxford University Press.
- , 2012b, “Which ‘key to all mythologies’ about the self? —A Note on Where the Illusions of Transcendence Come from and How to Resist Them”, *Immunity to Error through Misidentification: New essays*, edited by F. Recanati and S. Prosser, New York: Cambridge University Press.
- , 2017, “Stopping Points: ‘I’, Immunity and the Real Guarantee”, *Inquiry: An Interdisciplinary Journal of Philosophy*, 60 (3): 233–252.
- Descartes, R. , 1985, *The Philosophical Writings of Descartes*, edited by J. Cottingham and D. Murdoch, Cambridge: Cambridge University Press.
- Echeverri, S. , 2020, “Guarantee and Reflexivity”, *The Journal of Philosophy*, 117 (9): 473–500.
- Evans, G. , 1982, *The Varieties of Reference*, New York: Oxford University Press.
- Frith, C. D. , 1992, *The Cognitive Neuropsychology of Schizophrenia*, Hillsdale, NJ: Erlbaum.
- Gertler, B. , 2018, “Self-Knowledge and Rational Agency: A Defense of Empiricism”, *Philosophy and Phenomenological Research*, 96 (1): 91–109.

———, 2022, “Agency and Self – Knowledge”, *The Routledge Handbook of Philosophy of Agency*, edited by L. Ferrero, New York: Routledge.

Howell, R. , 2001, “Kant, the ‘I Think’, and Self – Awareness”, *Kant’s Legacy*, edited by P. Cicovacki. Rochester, NY: The University of Rochester Press.

Howell, R. J. , 2007, “Immunity to Error and Subjectivity”, *Canadian Journal of Philosophy*, 37 (4): 581 – 604.

Liang, C. , 2015, “Self – as – Subject and Experiential Ownership”, *Open MIND*, 24 (T), Frankfurt am Main: MIND Group.

Korsgaard, C. M. , 2009, “The Activity of Reason”, *Proceedings and Addresses of the American Philosophical Association*, 83 (2): 23 – 43.

Moran, R. , 2001, *Authority and Estrangement: An Essay on Self – Knowledge*, Princeton: Princeton University Press.

Palmira, M. , forthcoming, “Questions of Reference and the Reflexivity of First – Person Thought”, *The Journal of Philosophy*.

Perry, J. , 1997, “Indexicals and Demonstratives”, *A Companion to the Philosophy of Language*, edited by B. Hale and C. Wright, Oxford: Blackwell.

Pryor, J. , 1999, “Immunity to Error Through Misidentification”, *Philosophical Topics*, 26 (1 & 2): 271 – 304.

Rovane, C. A. , 1987, “The Epistemology of First – Person Reference”, *The Journal of Philosophy*, 84 (3): 147 – 167.

Shoemaker, S. , 1968, “Self – Reference and Self – Awareness”, *The Journal of Philosophy*, 65 (19): 555 – 567.

———, 1970, “Persons and Their Pasts”, *American Philosophical Quarterly*, 7 (4): 269 – 285.

———, 1994, “Self – Knowledge and ‘Inner Sense’”, *Philosophy and Phenomenological Research*, 54 (2): 249 – 314.

Wittgenstein, L. , 1958, *The Blue and Brown Books*, Oxford: Blackwell.