

马克思主义及哲学研究

# “信念之谜”与合理性的不确定性\*

陈常桑

**【提要】**对“信念之谜”的解答依赖于去引号原则和信念融贯性原则，而去引号原则受到主体信念之审慎性和真诚性的双重约束，可能出现的不审慎或不真诚的信念态度与不融贯的信念相结合，会导致“信念之谜”的难解。如果说主体信念受到合理性的信念融贯性原则约束，那么信念语境中的语句意义和信念指示词的指称会出现多重的不确定性，融贯性原则的这种不确定导致了合理性的不确定性。然而，可以运用维特根斯坦的“遵守规则悖论”和戴维森的合理性概念，对合理性的这种不确定性进行反向约束，从而保证合理性在某种程度上的确定性。

**【关键词】**信念 合理性 融贯性 不确定性

**【中图分类号】**B15 **【文献标识码】**A **【文章编号】**1000-2952(2012)04-0005-07

受“弗雷格之谜”(Frege's Puzzle)的启发，克里普克(Saul Kripke)把“信念之谜”(Puzzle of belief)当作真正的问题提了出来，这个“信念之谜”不仅涉及同一替换律、信念、意向性等重要理论，还涉及语言学、心理学、逻辑学等重要领域。然而，除了坚持名称的因果理论和批判描述理论，克里普克本人并没有为“信念之谜”提供令人满意的解决方案。在弗雷格主义的解决方案之后，哲学家们给出了直接指称理论、关联主义等解决方案。本文并不打算正面考察这些解决方案，而是绕到该谜题的一个侧面，从分析“信念之谜”与当代合理性理论及不确定性理论之间的密切联系入手，试图厘清“信念之谜”与人之合理性的哲学特质。

## 一、“信念之谜”的两个原则

在电影《超人》中，女主角露易丝在某报社担任记者，她的同事肯特是个人行不久的呆头呆脑的年轻人。露易丝不知道肯特就是近段时间外界风闻的在城市上空飞来飞去惩恶扬善的超人，她压根就没有把平凡无奇的肯特与所向披靡的超级飞人联系起来。当然，对于超人为什么要向心爱的姑娘隐瞒自己的身份，这似乎是个难解之谜。不过，无论如何，在这个故事中，以下两个句子对于露易丝来说都是真的：

- 1) 肯特不会飞。
- 2) 超人会飞。<sup>①</sup>

为了将上述句子与露易丝的信念语境联系起来，有必要引入一个看来非常自然的原则。在《一个信念之谜》<sup>②</sup>一文中，克里普克称之为“去引号原则”(the Principle of Disquotation)：

PD) 对于正常的说话者S来说，如果S审慎而又真诚地断定句子“p”，那么S相信“p”(其中替换“p”的句子不含索引、代词或者歧义等这些直观上违反原则的成分)。

\* 本文受上海社会科学院课题“跨文化沟通的分析哲学基础”项目资助，谨致谢忱！

① 黄敏：《分析哲学导论》，中山大学出版社2009年版，第280~283页。

② Saul Kripke, "A Puzzle about Belief", in *Meaning and Use—Papers Presented at the Second Jerusalem Philosophical Encounter*, Avishai Margalit (ed), D. Reidel Publishing Company, 1976, pp. 239-256.

这个原则在说话人对一个句子的心智态度亦即命题态度与说话人的信念内容之间建立了一种对应关系,使得我们可以据此认为说话人所断定的句子描述了他或她的信念内容。把这个原则运用于上述故事中,将1)和2)分别代入“p”,就得到下面两个信念描述:

- 3) 露易丝相信肯特不会飞。
- 4) 露易丝相信超人会飞。

露易丝并不知道“肯特”和“超人”所指的是同一个对象,因此即使3)和4)都可以合理地描述露易丝的信念状态,也不妨碍我们认为这两个信念描述相互矛盾。但是,我们是站在已经理解信念真相的角度做出如此断言的,然而这样一来,这个断言就在某种程度上掩盖了信念在本质上的某种难解之处。如果不假定这种理解,那么3)和4)的不相容关系就为“什么是信念?”这个问题投下了阴影——我们可以按什么方式把这两个不相容的信念解释得相容呢?受逻辑上的不矛盾律制约,主体通常并不准备接受两个互相矛盾的信念,例如我们不能同时认为一个东西既是白的又不是白的。但在信念语境中,不矛盾律似乎不起作用,我们可以把关于超人/肯特的两个相互矛盾的信念都归属于露易丝,这说明信念具有某种特异之处。这种特异之处被克里普克称作“信念之谜”。信念之谜的关键之点是,信念主体视角的两个相互融贯的信念,为何实际上可能是相互矛盾的。我们可将之归结为信念主体在认识论上处于信念失真状态,或者主体视角与客体视角之间的“断裂”,但问题不止于此。

#### (一) 去引号原则的两个限制性条件

根据克里普克的描述,去引号原则有两个必要的限制性条件:其一,审慎性条件。“审慎的”(on reflection)限制了这样一种可能性:说话者可能没有意识到他所说语词的意义以及偶然的概念上或语言上的混淆。其二,真诚性条件。“真诚地”(sincerely)意味着排除了主体信念的虚伪、反讽等类似的东西。克里普克之所以加上这两个限制,是想通过简单化的处理来看到信念之谜的本质。然而,实际生活中的信念状态往往并非如此简单。如果除去这两层限制,那么就存在如下两种可能性:

##### 1. 主体信念态度有时是不够审慎的

克里普克为自己的因果命名理论举了一个例子,以此来表明关于专名的因果链条模式不是一个理论,因为它并没有给出专名成功地指称其对象的充分必要条件。我们发现,这个例子同样可以用来表明说话者的信念态

度有时是不够审慎的。这个例子是:在数学课上,老师对学生说,最早求出圆面积的人名叫史密斯。史密斯其实是老师的一位邻居,但老师仅仅不想特意说明是谁最早求出圆面积,于是随口说出了“史密斯”这个名字。在这个例子中,可归属于老师的信念“史密斯最早求出了圆面积”基于不够审慎的信念态度之上,他对此信念的表达显得过于随意——当然也不排除他在此运用反讽手法的可能性。

##### 2. 主体信念态度有时是不够真诚的

说话者有可能是说谎者。我们不难设想说话者实际持有信念与他人根据其言语表达而归属于说话者信念之间的不融贯情形,这就是信念的虚假和伪装状态。本文要追问的不是说话者的信念态度未必是真诚的,而是判别说话者信念态度真诚性的标准或根据何在。后文将回到这点。

主体的审慎态度和真诚态度对信念归属(blief ascription)至关重要,如果没有这两条原则的约束,归属于主体的信念往往并不是主体的真实信念。对于主体的A和B两个信念来说,要判定其是否融贯,首先要取决于信念持有者的审慎态度和真诚态度。由于审慎态度和真诚态度的缺乏总是存在的,所以对于主体信念的外在判断是不确定的。

#### (二) 信念的融贯性原则

我们知道,去引号原则保证了在露易丝看来的“肯特不会飞”和“超人会飞”分别成为可归属于露易丝的两个信念3)和4)。然而,关于露易丝的信念之谜之所以能够产生,还需要具备另一个前提,即信念融贯性原则(principle of coherence)。信念的融贯性是人之合理性(rationality of human being)的逻辑要求,也是理解得以可能的“信念之网”的整体论要求。去引号原则是一个关于信念归属的直觉性原则,它涉及到我们如何理解信念,信念的融贯性原则也是如此。当肯特和超人指称相同时,从信念主体视角来看,露易丝的信念3)和4)仍然是融贯的,因而并未出现非理性情形。这两个信念之所以融贯,是由于露易丝的以下信念:

- 5) 露易丝相信肯特不是超人。

信念之谜的实质是,两个实际上相互矛盾的信念1)和2),为何在当事人露易丝看来反而并不矛盾。露易丝之所以能同时持有两个实际上并不融贯的信念3)和4),正是因为她缺乏对肯特与超人的同一性关系的认知。我们之所以认为3)和4)把两个互相矛盾的信念归属于露易丝,是因为我们把这两个信念看作是关于同

一个对象的信念，这种站在客体性视角的信念是从物信念 (*de re* belief)，其标志是指示词位于命题态度词的辖域之外，如这样一个信念：

6) 对于那个最矮的间谍，马利相信他是间谍。

然而，就3)和4)是通过去引号而得出的而言，又应当被理解为从言信念 (*de dicto* belief)，其标志是命题态度词项的辖域包含了整个命题，在其辖域之外并没有属于该命题的指示词，比如下列信念：

7) 马利断定“那个最矮的间谍是间谍”为真。

8) 马利相信那个最矮的间谍是间谍。

按照去引号原则，7)和8)具有真值可替换性。由此可见，从言信念与从物信念并不等价，当从言信念7)和8)总是为真时，如果事实上马利并没有认为那个最矮的间谍是间谍，那么从物信念6)则为假。若当作从物信念理解，则露易丝的信念3)和4)不融贯，因而导致非理性情形。可见，只要将信念区分为从言信念与从物信念，在去引号原则的限制之下，就可以更好地处理信念之谜所导致的信念融贯性问题。

从言信念是从主体视角出发的，因而同属于合理性和信念归属的信念融贯原则是一种内在原则。若果如此，要判定主体信念A和B是否融贯，缺乏一个外在的、客观的、公共可见的判断标准。要判断A与B这两个信念是否融贯，取决于对主体信念A和B的证据的考察，这是一个外在的和公共可见的工作。

不融贯的信念首先是非理性信念，对信念归属产生了很大的烦扰。克里普克在《一个信念之谜》首段说，对于信念之谜，任何解答都徒劳无益。但我们仍然有理由认为，从言信念与从物信念之区分有助于我们为信念之谜提供解释。在露易丝的故事中，如果露易丝的信念3)和4)相互矛盾，那是我们将之视为从物信念。然而，合理性所要求的信念的内部融贯，是从言信念的融贯。由从言信念到从物信念的过渡涉及到信念的归属问题，这个过渡往往会导致信念相互矛盾，这种矛盾继而导致合理性的信念融贯要求受到制约。去引号原则是信念归属的需要，信念的不融贯并不出现在信念归属阶段。如果这种不融贯出现在主体信念内部，则违反了合理性的信念内在融贯原则；如果这种不融贯出现在主体视角的从言信念与客体视角的从物信念之间，则出现了信念之谜。

## 二、合理性的信念融贯性原则

信念之谜对合理性提出的问题是：为什么说话者的

两个理性的信念实际上仍然可能是相互矛盾的？下面让我们转到对信念的合理性维度的讨论。

在《行动的合理性》一书中，塞尔 (John Searle) 认为，合理性的理论传统是沿着亚里士多德到休谟再到康德这条线下来的。显然，这个传统并非铁板一块，而是在不同哲学家那里各有侧重，但这并不妨碍塞尔把这条主线称作合理性理论的“经典模型” (Classical Model)。他还指出，休谟为这个模型提供了最清楚描述，并进而给出了合理性的经典模型的六个设定：

1. 理性的行动由信念和愿望所致。
2. 合理性是关于遵守规则的问题，即遵守那些区分出理性思想行为与非理性思想行为规则。
3. 合理性是一种分离的 (separated) 认知能力。
4. 只有在行动的心理前件中发生错误时，才能发现意志薄弱 (古希腊人称之为 *akrasia*) 的明显情形。
5. 实践理性从一个行动者的主要目标的清单开始，它包括这个行动者的目的和基本愿望、目标和意图；这个清单并非因为其自身而受到合理性的约束。
6. 仅当主要愿望系列处于一致状态时，才能出现一个合理性的整体系统。

塞尔认为，这六个设定并不在同一个哲学家那里全部具备，但它们构成了传统哲学家对合理性观念的一个内在融贯的整体。<sup>①</sup>

我们从近年出版的《牛津合理性手册》中可以找到合理性概念的当代视角：其基本内涵是对信念的内在融贯性和行动的逻辑一致性要求，其中涉及到“信念的形成”，旨在“达到真信念以及避免假信念”的合理性要求，被看作是理论合理性 (theoretical rationality)，涉及到“行动的表现”，旨在“按照某人关于他想得到什么以及如何得到的信念而得到他最想要的东西”，被看作是实践合理性 (practical rationality)。<sup>②</sup> 一般来说，信念是理论合理性的核心，当信念是真的并且能得到证成 (justification) 时，就构成了知识，知识因而成了理论

<sup>①</sup> John Searle, *Rationality in Action*, ch. 1, The MIT Press, 2003.

<sup>②</sup> 参见 Kirk Ludwig “Rationality, Language, and The Principle of Charity”, in *The Oxford Handbook of Rationality*, p. 345.

合理性的追求目标；愿望或偏好是实践合理性的核心，当愿望和偏好被确认并具备必要的外在条件时，就促使行动的完成，行动因而成了实践合理性的追求目标。合理性的两大原则：信念的融贯性原则和行动的逻辑一致性原则。这种信念融贯必须是内在融贯，即一种“在当事人看来”的融贯性。这种内在融贯性是一种从言融贯。

### （一）戴维森的合理性准则

综观戴维森的著作，他对“合理性”一词的使用与上面提到的从亚里士多德经休谟到康德的合理性线路和《牛津合理性手册》的基本精神是大体一致的：一言以蔽之，合理性问题主要关注信念与行动的网状系统内部的逻辑关系。在某种意义上，戴维森的合理性理论是这种“经典模型”的继承和改进。戴维森在合理性问题上的贡献不在于他为之提供了一个与众不同的定义，而是将合理性当作一个其地位仅次于绝对之真的基础概念，并与善意原则、整体论原则等方法论原则一道，参与建构了他的语言哲学、心灵哲学、价值哲学、知识论和行动哲学，并以一种别开生面的方式处理了非理性问题。戴维森的哲学中的合理性同样侧重于形式合理性；与罗尔斯、麦金泰尔等侧重于从实践领域把握合理性问题的哲学家不同，戴维森在不同方面同时兼顾到理论合理性和实践合理性问题。

合理性的一个要求是信念的内部融贯性，并且在戴维森看来，并非先有信念然后再谈信念是否融贯的问题，而是先具有某种程度的合理性（在此表现为融贯性）和一致性才会有信念的产生：

因为信念被它们与其他信念的关系个别化和确认这一事实，如果一个人要拥有一些信念，他必须拥有大量的信念。信念互相支持，彼此赋予对方以内容。信念之间因此具有逻辑关系。结果是，除非一个人的信念彼此之间高度地相容，否则信念的内容得不到确认。某种程度的合理性和相容性因而是拥有信念的一个条件。<sup>①</sup>

这种信念的内部融贯性其实就是信念的逻辑一致性。这表明说话者的信念要受到整体论的约束。

### （二）合理性的三重不确定性

克里普克的认识论论证缺乏从言信念到从物信念的过渡。这种过渡要依赖于说话者的两种信念态度：审慎态度和真诚态度。前文提到的去引号原则的审慎原则，类似于合理性中的“考虑周全原则”（principle of all-things-considered，简称ATC原则），这种判断是关于对一个人根据所有信念、愿望和其他态度而做出

最佳的行动选择。这种考虑周全的判断使各种理性活动结合在一起构成一个人的慎思，就是这些事情的结果：解决一个人信念上的决策，制定出一个人信念和其他态度的含义，以一种特定的秩序来为自己的偏好排列等级。所有这些理性活动都由符合一种特定规范性的合理性要求进行指导。信念的非融贯导致自我欺骗的出现。

因此，合理性的内在融贯原则具有某种不确定性，这与彻底解释的不确定性是对应的，它们都依赖于说话者的持有信念或说话言语的审慎态度和真诚态度。就合理性原则来说，如果信念主体持某种非审慎和非真诚态度，那么对于其信念是否理性的判别就相当困难了。

由此可见，合理性有三个层次的不确定性：本体论的不确定性、认识论的不确定性和语义学的不确定性。本体论的不确定性是指，就信念是主体的命题态度而言，其内在融贯性总是处于一个不确定的状态；认识论的不确定性在于，就信念是主体信念来说，对主体信念之融贯性的客观的或“第三人称”视角的认识，往往是不确定的，因为它受到审慎性和真诚性因素的制约；语义学的不确定性是指，就信念是命题态度而言，命题的意义和语词的指称，都存在这样那样的不确定性。

## 三、信念与语义：意义和指称的不确定性

判别说话人的信念是否理性，不仅要看其是否内在融贯，以及主体信念态度是否真诚或善意，还要看信念的语义内容是否确定。如果我们对信念语句的意义和信念指示词的指称并不存在足够的确定性，那么合理性就不具备足够的确定性。下面探讨导致这种合理性的不确定性的意义因素和指称因素。

### （一）意义的不确定性

在一位戴维森式的哲学家看来，一种意义理论就是一种解释理论，而彻底解释（radical interpretation）为解释理论提供了一种有效的分析模式。在彻底解释中，其所依赖的方法论原则——善意原则是不确定的，这导致了彻底解释的不确定性。前文说到去引号原则中的真诚态度相当于信念解释中的“善意原则”（principle of charity）。

彻底解释存在双重约束，一个是基于约定的对于

<sup>①</sup> Davidson, "The Emergence of Thought", *Subjective, Intersubjective, Objective*, p. 125.

句子语义指派的形式约束，另一个是对于信念指派的经验约束。由于语义与信念的对应性，这种双重约束同时既是对语义指派的约束，也是对信念指派的约束。然而，即使在这种双重约束之下，彻底解释仍然是不确定的。同样，对于违反善意原则的情况，将对信念归属带来极大的麻烦，使得我们很难确定什么样的信念可合理地归属于说话者，遑论对其信念的内在融贯性作出判定了。

对于戴维森来说，彻底解释的不确定性并没有让戴维森感到迟疑不决。对于彻底解释的目的而言，即使得不到确定的意义理论，但意义这个概念却得到了阐明。意义理论的不确定性，无损于意义概念的确定性。彻底解释虽然不确定，但作为语言哲学分析模式，仍然是充分的和必要的。我们同样可以认为，意义和指称的不确定性也不应当使我们放弃对合理性的确定性寻求。即使得不到确定的合理性，但合理性这个概念却由此得到了阐明。

## (二) 指称的不确定性

仅以露易丝的“信念之谜”为例，“肯特”作为专名，其指称是不确定的。对此我们可以找到四个理由：

### 1. 蒯因的“指称的不确定性”理论

在《语词与对象》一书中，蒯因设定的田野语言学家通过“彻底翻译”学习土著部落语词的情景，确定土著什么时候表示赞同什么时候表示反对，并不是一件容易之事：首先，土著不一定有两个明确的说法分别表示同意/不同意；其次，即使有这两个词，还要分辨哪个对应同意，哪个对应不同意；最后，土著未必愿意配合语言学家，他可能说谎、说错，诸如此类。这里遇到的不是学习过程中的某种暂时性障碍，而是原则上不能澄清的东西，蒯因称之为指称的不确定性（indeterminability of reference）。

### 2. 维特根斯坦的“不定簇理论”

弗雷格—罗素主张专名等于缩略的描述语，这个主张面临一系列问题：一个专名等于一个、一些还是全部描述语？若是等同于一个或一些，则等同于哪个或哪些？对每一个人是否一视同仁？若是等于全部，若我们发现了专名指称之事物的新性状怎么办？在《哲学研究》中，维特根斯坦修正了他早期的名称理论，同时也对罗素的摹状词理论提出了自己的新见解。<sup>①</sup> 罗素认为，“摩西”这个名称可以由各种各样的描述来定义，例如定义为“那个带领以色列人走过荒漠的人”，“那个生活在彼时彼地、当时名叫‘摩西’的人”，“那个童年时被法老的女儿从尼罗河救出的人”，等等。然而问题是，我们假定这一个或那一个定义，“摩西没有存在过”这个命题就会有不同的意思。这导致一个结论，如果我们

要用意义来谈论名称 N，就必须不在固定的含义上使用名称 N，并且这样并不削弱它的用途。因此，一个合用的名称，必定是若干描述的不确定的集合。

### 3. 克里普克反驳指称的描述理论的认识论论证（epistemological argument）

维特根斯坦和后来详尽发展出一种描述理论的哲学家塞尔，尽管与弗雷格—罗素的专名理论有诸多不同，但他们都认可后者的一些基本主张：专名是描述语的某缩略形式。比如塞尔主张一种关于专名的簇摹状词理论（the cluster description theory of proper names）。简单的描述理论可以表述为，专名 n 成功地指称对象，当且仅当存在摹状词 P，使得存在唯一的对象 o，P(o) 为真。一言以蔽之，专名是摹状词的缩写形式。塞尔为这个理论提供了一种精致版本，他把原来的那个唯一的摹状词 P 改成一簇数量不定的摹状词，只要对象 o 满足了这组摹状词中足够多的摹状词，n 就指称了 o。根据这种理论，比如对于“亚里士多德”这个专名来说，这组摹状词中就包括“柏拉图最知名的学生”、“亚历山大大帝的最博学的老师”、“写出《尼各马可伦理学》的那个哲学家”等等。一个对亚里士多德了解不多的中学生可能知道亚里士多德是柏拉图最知名的学生，也知道他还是亚历山大大帝最博学的老师，但即使他不知道《尼各马可伦理学》的作者是谁，他仍然知道“亚里士多德”这个专名的指称是什么。

克里普克的认识论要驳斥的观点是，专名的指称是通过摹状词所确定的。之所以称之为认识论论证，乃是由于它所驳斥的观点是以下两个观点的合取：（1）关于对象的指称以关于该对象的知识为前提；（2）这种知识是描述性的。这个观点并不承诺表述这种知识的命题必然为真，而只承诺其为真，因而就是这样一个观点，即对于任何专名 n，必然存在摹状词 F，使得 F(n)。如果对描述论放宽要求，我们可以表示如下：

PR) 专名 n 成功地指称了对象 o，仅当存在摹状词 F，o 满足 F。

克里普克设计了一个反事实情形来驳斥这个观点。关于哥德尔，假定我们所知道的仅仅是他证明了算术不完全性定律，于是我们用这个描述来确定“哥德尔”一词的指称，由此相信句子“哥德尔证明了算术不完全性定律”为真。然而，假定真实情况并非如此，算术不完全性定律不是哥德尔证明的，而是另有他人，这个人就是施密特。由于某种原因，这个定律

<sup>①</sup> Ludwig Wittgenstein, *Philosophical Investigations*, § 79.

被“张冠李戴”地归功于哥德尔，于是情况就成了证明不完全性定律的人是施密特。现在，按照描述理论，我们使用“证明算术不完全性定律的那个人”这个摹状词所确定的指称实际上就是施密特。这样一来，PR)中的前件是真的，因为专名“哥德尔”成功地指称了现实世界中的哥德尔，但后件却是假的，因为现实世界中的哥德尔并不满足其中的摹状词。克里普克的论证属于“误识别论证”(the argument from mis-identification)，通过此论证，我们可以看到专名与摹状词在构成知识内容方面的不同作用。克里普克这样总结道：“如果哥德尔的欺骗行为大白于天下，那么哥德尔就不会再被称作‘不完全性定律的发现者’了，但他仍然会被称作‘哥德尔’。因此专名不是摹状词的缩写。”<sup>①</sup> 克里普克的结论是：专名的使用并不以认知情况为前提，专名的指称不是通过描述性的内容或者摹状词所确定的，换言之，专名并没有认知内容。然而，立即出现了两个问题：(1) 专名的使用毕竟要以指称已经确定为前提，如果专名的指称不是通过描述的方式确定，那么又是以何种方式得以确定的呢？(2) 如果专名并无认知内容可言，那么如何解释通过专名可以构成类似于“晨星=昏星”这样的具有认知意义的命题呢？对问题(1)的回答，形成了《命名与必然性》中的著名的因果命名理论，克里普克通过给出命名的因果链条模式(causal chain mode of naming)，大体上做出了回答。而问题(2)迫使克里普克对“弗雷格之谜”(Frege's puzzle)<sup>②</sup> 作出自己的回答。他故而论证说，解释此类命题的认知意义，并不属于语义学的任务。

#### 4. 对克里普克因果命名理论的反论证

饶有趣味的是，在同一本书里，克里普克对自己的因果命名理论提出了反论证。在上文我们说道，因果命名理论并不是一个严格的理论，因为它并没有为专名成功地指称其对象提供充分必要条件。同样在上文所借用的克里普克为其因果命名理论所举的反例中，学生关于史密斯所知道的东西不能被归于老师的那个邻居。虽然学生使用“史密斯”这个名字在因果链条上来自于老师的使用，但这一点并不意味着学生使用这个名字所指的就是老师的那位邻居。就这一点而言，老师使用的并不是学生确定这个名字的指称的充分条件。

## 四、对合理性之不确定性的反向约束

如果信念的融贯性是一种内在融贯性，判别融贯性的标准也是内在的，并不存在一个公共可见的信念融贯

性标准，那么就会出现这样的问题：信念持有者是如何遵守只属于其个人内在融贯性标准的呢？这个问题把我们引致维特根斯坦关于遵守规则的讨论。

### (一) 维特根斯坦：并不存在私人的融贯性标准

在《维特根斯坦论规则和私人语言》中，克里普克认为维特根斯坦在《哲学研究》中给出了一个关于遵守规则的悖论(即“规则悖论”)，而这个悖论导致了关于意义的怀疑论。也就是说，我们没有依据构成关于意义的断定，换句话说，不存在关于意义的事实。克里普克进而认为，如果说语言是一种规则系统，那么规则悖论是私人语言论证的核心所在。在他看来，只有在语言共同体中才谈得上意义，这就直接否定了存在私人语言的可能性。他这样说未必妥当，但有一点是确定的，就是维特根斯坦试图通过规则悖论提示人们，并不存在私底下遵守规则的情况，因为个人无法为遵守规则或违反规则提供确切的标准，正是这一点，为主体信念的内在融贯性的不确定性提供了某种反向约束。

### (二) 戴维森：“合理性”是一个规范性概念

合理性有个程度问题，人类是一种理性不完满的生物，一个理性人也只是在某种程度上是理性的。人之合理性是大致的、有限的合理性。在戴维森看来，说合理性有个程度问题，亦即不是说任何人在任何情况下都是完全理性的；实际上，任何人都可能在逻辑上犯错误，或者同时持有彼此不一致的信念。但是一个人无论如何不一致，他仍然有一个限度，即便是不一致性或非理性的其他形式，也只能在理性的范围之内出现；不一致性是合理性的烦乱(perturbation)，而不是合理性的缺失；合理性的程度并不是指不同生物在合理性方面处于不同的阶段，而是指不同的人以及同一个人不同的状态下在合理性方面的不同程度。行动者理解世界的条件并不依赖于解释者对被解释者方面发现完美的合理性，而是“足够多的合理性”。按照普特南的说法，理性生物

① Saul Kripke, *Naming and Necessity*, Harvard University Press, 1980.

② 弗雷格在“Sense and Reference”一文中指出，对于“1) 晨星=昏星”和“2) 晨星=晨星”两个句子，前者表达了一个天文学上的认知发现，而后者却没有，其缘故究竟何在，这通常被称为“弗雷格之谜”。他对此问题的处理可以这样表述：1) 和 2) 之所以有区别，是因为它们表达的不是同一个命题。如果 1) 是真的，那么它与 2) 具有同样的真值，即在意谓上相同，亦即“晨星”和“昏星”的意谓相同，但这并不意味着二者表达了同样的认知意义。克里普克认为，“弗雷格之谜”并不是一个真正的问题，因为它可以归结为信念之谜。

之话语在合理性方面有“理性的 (rational)”、“半理性的 (semi-rational)”和“离理性不远以至多少仍可理解的” (not-too-far-from-rational-to-still-be-somehow-intelligible) 三个不同的阶段。<sup>①</sup>

此外，说合理性是个程度问题，并不意指我们普遍来说都处于一个相当低的理性程度，或者人之非理性的情况要比理性的情况多得多。相反，在大多数情况下，我们都是理性行事的。其实，也只有这样，说我们有时是非理性的，才有意义。正如戴维森在《非融贯与非理性》一文中所指出的：“一个行动者不可能不在大多数时间与合理性的基本准则相一致。这个事实使非理性成为可能。”<sup>②</sup>被解释者认为真的大多数（更不用说所有）信念都是假的，这不符合理性的规范性要求。认为人总是犯错误，或从不犯错误，这些说法都不符合合理性的程度性要求。也许可以这么说，人有理性，这是合理性在绝对性方面的要求；人有非理性，这是合理性在程度上的要求。

信念归属活动类似于彻底解释活动。我们知道，信念归属的去引号原则所需要的两个下属原则——审慎性原则和真诚性原则，如果得不到保证，就可能导致信念归属的不确定性。戴维森的彻底解释理论也会遇到这个问题，于是他引入了另外两个原则——“考虑周全原则”和“善意原则”来分别对之进行制约。这两个原则分别对信念归属的双方——信念主体和信念归属者，以及彻底解释的双方——说话者和解释者的态度给出了限制。“考虑周全原则”和“善意原则”都是合理性之确定性的必要条件。

### (三) 小结

判别说话人的信念是否理性，不仅要看看其是否内在

融贯，以及主体信念态度是否真诚或善意，还要看信念的语义内容是否确定。如果我们对信念语句的意义和信念指示词的指称并不存在足够的确定性，那么合理性就不具备足够的确定性。如果信念的融贯性受到了威胁，那么可能陷入“非理性悖论”的不仅是理论合理性，还可能包括实践合理性，因为信念与信念的不融贯可能蕴涵着信念与欲望的不融贯。

然而，可以运用维特根斯坦的“规则悖论”和戴维森的合理性概念，对合理性的这种不确定性进行反向约束，从而保证合理性在某种程度上的确定性。合理性的不确定性要受到两方面的约束。其一，规则悖论使得私底下遵守融贯性标准并没有意义。信念的融贯性一方面是内在的，只有信念持有者本人才能断定信念 P，并进而断定 P 与其他信念（甚至还有欲望等非信念）是否融贯；另一方面，说信念持有者私底下遵守融贯性标准是毫无意义的，实际上他遵守的还是外在标准。其二，合理性作为一个规范性概念，它在相当大程度上是对理性存在者的一个强制性要求。

本文作者：复旦大学哲学博士后流动站在站博士后，上海社会科学院哲学研究所助理研究员，中国社会科学院研究生院哲学系 2010 届博士  
责任编辑：周勤勤

① See Hilary Putnam, *Representation and Reality*, The MIT Press, 1991, p. 87.

② Davidson, “Incoherence and Irrationality”, *Problems of Rationality*, Oxford University Press, 2004, p. 197.

## “Puzzle about Belief” and Indeterminacy of Rationality

Chen Changshen

**Abstract:** It depends on “principle of disquotation” and “principle of coherence” to resolve the “puzzle about belief”, while “principle of disquotation” is constrained by prudence and sincerity of belief, and the possible imprudence or insincerity combining with incoherent belief would make the “puzzle about belief” confusing. If the subject’s belief is constrained by principle of coherence of rationality, then the meaning of sentences and reference of belief-referring words in the context of belief would be faced with multiple uncertainties, and the uncertainty of “principle of coherence” will cause the indeterminacy of rationality. However, we can use Wittgenstein’s “rule-following paradox” and Davidsonian rationality, providing reverse constraints for these indeterminacies, thus ensuring the determinacy of rationality in some degree.

**Key words:** belief; rationality; coherence; uncertainty