逻辑学研究　2026 年第 1 期，149–165
文章编号：1674-3202(2026)-01-0149-17

密涅瓦猫头鹰问题的困境与自然规范性出路

蔡灵新

摘　要：谬误是非形式逻辑中用于研究论证的重要部分。然而，艾金在他的演讲中提出一个担忧：作为元语言的谬误一旦被具有交互性特质的人类所拥有，则会引发关于谬误的谬误（fallacy fallacy），并在此基础上继续衍生出谬误的谬误的谬误，从而形成悲观的密涅瓦猫头鹰问题。戈登对这个问题进行解构并认为其根本原因并不是结构性问题，而是动机性问题，从乐观的角度看可以提供辅助规则修正来避免大部分误用。本文认为动机不是密涅瓦猫头鹰问题的根源，该问题的本质在于将忽略了论证中除了语言之外的信息，没有充分考虑论证规则的规范性来源。本文结合吉尔伯特的自然规范性理论讨论谬误使用的问题，并以此分析密涅瓦猫头鹰问题的局限性及其出路。

关键词：密涅瓦猫头鹰问题；病理性循环；元语言；谬误的谬误；自然规范性

中图分类号：B81　　**文献标识码：**A

1　引言：论证中的密涅瓦猫头鹰问题

密涅瓦猫头鹰问题（The Owl of Minerva Problem）是艾金（S. F. Aikin）在 2020 年的演讲中提出来的。这个问题的名称取自黑格尔（G.W.F. Hegel）在《法哲学》（Philosophy of Right）的序言中写到“密涅瓦猫头鹰只在夜色降临时才起飞”。密涅瓦猫头鹰是智慧女神的象征。这句话的意思是，智慧只有在一天结束的时候才会产生。黑格尔的这句话表达了智慧是经验的产物，是后验的，具有回溯性，因此难以履行指导未来行动的任务。艾金认同黑格尔的观点，并认为对于具有互动性的人类来说，这是一种难以避免的病理性循环（pathological loops）。

艾金根据事物与他者的互动情况，将事物区分为两类。一类是“表现方式与我们如何谈论或分类它们无关”（[5]，第 15 页）的，称为无关型类别（indifferent kinds），例如物理物质、微生物等。另一类则是“表现方式会随着我们如何分类或谈论它们而改变”（[5]，第 15 页），称为互动型类别（interactive kinds），例如人类。[1]具有互动性的人类会根据他们被他者讨论、评价、分类而发生改变。他们的改变也会反过来改变他者对他们的理解和评价。因此，会形成一个包括第一人称

收稿日期：2025-05-20
作者信息：蔡灵新　中山大学哲学系
lx_cai@163.com

[1]艾金关于无关型类别和互动型类别的命名引用自哈金，参见 [15]。

动机、慎思和行动、第三人称的反思判断的信息循环。（[14]）即，理解对象会被理解的信息所改变。因此，“对于互动型类别来说，我们在寻求对象的过程中，也在把对象推的越来越远”。（[5]，第 16 页）换言之，当对一个人的行为作出指导时，是对已经作出的行为作出的指导。这个指导可能会影响行为者接下来的行为，但由于接下来的行为是未知的，对新的行为的指导也只能够是在新的行为出现之后才能产生。因此，这种指导是回溯性的，即只能够对之前的行为作出理解和总结，而不能够确切地指导接下来的行为。

艾金悲观地认为，这种回溯性不仅仅无法对接下来的行动进行指导，还会产生病理性的循环。他指出，在非形式逻辑中最令人的担忧的循环是谬误的谬误（fallacy fallacy）。谬误（fallacy）指的是一种看起来有说服力、但其实推理有漏洞、或者论证结构不正确的论证模式。本文讨论的论证理论中的谬误是非形式谬误，涉及对论点内容、例证或语言运用的问题。谬误是用来评价论证的元语言（metalinguistic）。不同学者对谬误的谬误的定义不同。阿伯丁（Andrew Aberdein，[1]）将对谬误的谬误的定义分为两种：第一种认为谬误的谬误是对谬误的错误识别；（[9, 17, 18]）第二种认为谬误的谬误是指当某人在对话中指出对方论证中的谬误时，以此为理由否定对方的结论。（[5, 10]）艾金所讨论的谬误的谬误是第二种。它的基本形式是：

论证 A 支持命题 P，
A 中存在谬误，
因此命题 P 是错误的。

由于很多正确的命题 P 可能被错误的理由支持，因此谬误的谬误成为一种新的谬误。艾金发现，只有当某人拥有“谬误”的概念的时候，才会产生谬误的谬误。虽然谬误列表的提出是为了消除谬误，但实际上因此创造出了新的谬误。而随着谬误的谬误的产生，又可能继续产生谬误的谬误的谬误、谬误的谬误的谬误的谬误……“我们用来纠正的概念，实际上会带来需要纠正的新事物。”（[5]，第 18 页）因此，这是一种结构性的病理循环。虽然论证无效并不蕴涵结论为假是逻辑学的基本常识，但是，即便是具备良好逻辑素养的人在激烈的辩论实践中也可能会犯此错误。这种看似简单的逻辑诊断，在实践中经常会失效。事实上，不管是第一种谬误的谬误，还是第二种谬误的谬误，都会产生这样的病理循环。

如果谬误列表的产生为这种病理性循环的形成提供了可能性，且这样的循环无法停止，那么是否不讨论谬误，反而会更好地帮助论证者接近论证的真相？如果是这样，那么人类似乎应该停止所有的反思性行为、停止规则的制定、停止教育，回归原始。这显然不可能。那么产生这种谬误的病理性循环的原因是什么？这种病理性循环是否可以被解开？本文在接下来将针对这两个问题展开讨论。在文章的第二部分，本文将重构大卫 • 戈登（D. Godden）对这个问题的较为乐观的分

析，他认为，密涅瓦猫头鹰问题是动机问题而不是结构性问题，除去动机的问题所产生的悲观后果，这样的循环可以通过积极/正面的元语言指导新的论证。本文的第三部分将进一步讨论动机在这个问题中的作用，并尝试说明动机并不是产生这个问题的根本原因。在第四部分，本文将引入迈克尔·吉尔伯特（M. A. Gilbert）的自然规范性概念，论证病理性循环是产生于对规范性的误解。最后，本文回到密涅瓦猫头鹰问题，探索通过自然规范性产生的对密涅瓦猫头鹰问题限制。

2　戈登：这是一个动机性问题

在艾金的论述中，他认为元语言就是产生于经验的智慧，而元语言是产生关于论证理论的病理循环的根源。戈登（[14]）在此基础上，通过分析元语言（meta-language）和对象语言（object-language）之间难以被区分的关系，认为密涅瓦猫头鹰问题不是元语言的结构性问题，而是动机问题。根据他的观点，在没有动机问题的情况下，元语言所引发的误解和错误是可以通过话语引导和修正技巧得到改善的。

戈登分别定义了元语言和对象语言，认为两者虽然不同，但却难以完全分离。元语言是指分析和评价某一内容，表达对内容的态度或观点的语言。例如，谬误就是用来评价论证内容的元语言，谬误的谬误就是用来评价谬误的元语言。对象语言是指实际用于表达、描述、陈述对象世界的语言。例如，论证者所提出的论据。首先，一句话可能既是元语言，也是对象语言。举个例子：

论证者 A：这场比赛一定会赢，因为专家说了会赢。
论证者 B：你犯了诉诸权威谬误，未到比赛终结不能下定论。

在这段对话中“你犯了诉诸权威谬误”既是元语言，也是对象语言。当论证者指出对方的论证中存在谬误，如果他只是对对方的论证作出评价，那么这是元语言；如果他是用指出谬误的方式进行进一步的论证，这时候指出谬误也可以是对象语言。其次，论证需要大量的元语言知识。学习如何评价一个论证是学习如何进行论证必不可少的一部分。论证者需要先知道什么样的论证是好的，或者是符合标准的，之后才能够进行合理的论证。缺乏这种学习，则可能会陷入非理性的争吵。再则，论证本身就包含了“内在”的元语言元素。当论证者提出一个论证时，实际上暗示了他认为这个理由是支持结论的，他认为的这个论证是好的。最后，用元语言评价论证是同时涉及对对象语言和元语言层面的评价的。对于对象语言的评价是考虑对象语言的真值，对元层面的评价则是考虑接受这个前提的标准是否合理。综上所述，戈登认为元语言和对象语言难以明确分离，但这并不意味着元语言的出现必然伴随着病理性循环的产生，而是对元语言的公正性的破坏才会导致病理性循环。

戈登认为对元语言的公正性的破坏可能存在于两种情境。第一种情境是当元语言被用来支持论证者的信念偏见（belief bias）时。信念偏见是指当论证者持有某种信念时，认为自己持有的信念是绝对正确的，认为自己有权使用元语言来维护现有的观点，而不是通过讨论来调整认知自我。这时候，当行为者使用元语言对论证进行判断时，他们会对赞同的观点进行肯定，对不赞同的观点进行否定。例如，当一个人持有不能说谎的信念偏见，他会赞扬诚实的人，认为说实话是勇敢的、好的；而对于说谎的人，他会认为这是不理智的、虚伪的。在这个使用元语言的过程中，他并不会去考虑具体发生了什么：诚实是否会带来更大的伤害，或者说谎是否可能是一种更为合理的举措。元语言在这里是被用来强化持有信念者的认知偏见。第二种情境是在对抗性论证中，论证者只想要赢的时候。戈登指出，当论证者只想要赢的时候，不想承认自己的失败的时候，作弊的诱惑就会增强。这两种情境都是基于论证者动机，一方面是认为自己是对的，另一方面则是即便自己不对，也认为自己要赢得论证。

戈登对这两种元语言被误用情境进行进一步细分，认为造成这两种元语言误用的失误可能是意外的（accidental），也可能是出于算计的（cynical）。意外的失误是指论证者本身没意识到自己是在对元语言产生误用，即没有意识到自己产生了信念偏见或者被想赢的欲望所蒙蔽。这种对元语言的误用可以通过提供辅助规则来进行修正。例如，对误用者进行教育，让其发现自身的失误来纠正。出于算计的失误是指论证者故意将此作为策略，用来实现对自己信念偏见的巩固或赢得论证。戈登认为出于算计的失误是论证者的价值问题，是难以纠正的：

> ……值得考虑的是，我们无法制定一项规则，使参与者都重视那些在我们的实践中体现的价值和利益。或者更确切地说，虽然我们可以制定这样的规则，但它将完全无效。那些已经重视这些价值和利益的人，会按照已经存在的规则行事，并以这些规则的精神为指导。相反，那些不重视我们实践中所体现价值的人，在面对一条恳请、强制或要求他们内化这些价值的规则时，不会获得任何额外的动力。 （[14]，第 47 页）

戈登认为，每个人都会根据他们认同的规则来自行管理自己。这种认同感是每个人产生动机的基础，无法受到外部强加的规则的撼动。对于这种来源于价值的差别所产生的对元语言的“误用”，戈登认为是无解的。“我们只会要求自己对那些我们认为是我们自己的规范负责。”（[14]，第 53 页）密涅瓦猫头鹰问题的深层根源在于每个人对理性规范、标准、价值和目标的理解和认同。这并不是规范设立本身的问题。

戈登认为虽然价值问题难以调和，但并不意味着反思性的元语言只会带来病理性循环。如果仅仅将元语言的作用解读为避免在接下来的行动中犯错，那么元

语言确实可能会导致更多的错误，从而形成像谬误的谬误的病理性循环。但是，元论证语言不仅仅是用来批判和指出错误，它还具有分析性（帮助我们更清楚理解自己在争论中在做什么）和评估性（用来表扬成功的辩论技巧）。也就是说，根据戈登的观点，我们可以用这些元语言来赞扬某些优秀的辩论行为，激励人们追求更好的表现，从而形成一种积极的、向上的循环。例如“辩证的优雅”（dialectical elegance）的评价性元论证概念，它就可以作为一种赞赏的术语使用，从而产生能激励和鼓舞，让那些作为概念响应者的理性者去追求同样值得称道的成就。

戈登的论证尝试用价值的无法统一性来解释元语言所带来的病理性循环，并称之为动机问题。但是，拥有相同价值的人就会遵循同样的规则吗？或者价值不同的人就不会遵循同样的规则吗？动机对规则的误用的影响似乎仅仅是元语言的病理性循环产生的原因之一。正如伦理学中经常讨论的“好心办坏事”的问题一样，元语言能够带来好的作用或者坏的作用除了动机问题之外，似乎还需要考虑到更多的内容，例如目的、语境、参与者等等。戈登最后提出的对元语言的功能的乐观分析，恰巧说明了元语言作为论证工具，本身就存在被论证者好地应用，或者坏地应用的可能。而这些好与坏的定义，并非是否遵循规则能够说明的。

3　动机不是问题的根源

根据上一部分的分析，戈登认为元语言虽然与对象语言具有语义距离，但由于元语言与对象语言二者难以完全分离，因此容易出错。正如艾金和塔利斯所说，“当元语言的概念被用作持续争论中的一阶工具时，元语言的公正性就会被破坏。”（[8]，第 182–183 页）当元语言被误用时，是被作为“争论中的一阶工具”，即一种对象语言来使用，这时候的元语言相当于某种规则。而戈登认为这种错误并不一定会产生病理性循环，也可能产生好的循环。产生病理性循环的问题的根源在于论证者的动机，即论证者是否认同元语言所对应的规则的价值。那么，戈登的论证可以简单总结如下：

> 只有当（a）论证者将元语言误用为对象语言，并将其作为论证的工具（结构性问题），且（b）论证者不认同规则在实践中体现的价值（动机性问题），那么（c）会产生病理性循环。

根据戈登的观点，（a）和（b）是实现（c）的充分必要条件。换而言之，论证者以不同的价值将元语言误用为对象语言，则会产生病理性循环。只有两者同时实现，才能够使元语言病理性循环问题的发生。然而，这是存在问题的。

首先，这里提到“不认同规则在实践中体现的价值”，需要先解释什么是论证者应该认同的“规则在实践中体现的价值”。根据戈登的论述，似乎存在一种论证的普遍的价值。他认为，论证的普遍价值是论证实践所追求的积极目标，如真理、

理性自我完善、自主性、民主公正、以及通过自我调节提升论证质量等。在论证中，这种价值是一种自我修正与理性协商的正向价值。而违背这种普遍价值就是将元语言武器化（weaponization），把元语言当作工具来获得权力或压制对手，从而破坏论证实践的规范性与自我调节能力。根据戈登的论述，论证的价值是二元的，要么是好的，要么是坏的。实际上，这个想法过于理想化。在日常论证中，论证者几乎很少会完全持有一个积极的价值观，或者完全消极的价值观。因此，论证者可能在实践中有部分地追求自我完善，但同时将元语言当作武器来赢得论证。

考虑艾金（[5]）关于学生的例子，学生在学习了谬误理论之后，在论证中制造了谬误的谬误。在这个例子中，学生在学习时是认同谬误理论的价值是帮助他追求真理。当他在学习之后将谬误“武器化”时，可能是为了赢，也可能只是为了展现他的聪明，也可能只是为了表现他对谬误的活学活用。不管当下他认为的谬误理论的价值是什么，都不一定会否定他认同谬误的价值是让他能够在论证中追求真理。在这个过程中，学生所认同的价值不是纯粹的，他既认同一种正向的价值，同时也认同一些与正向价值不同的东西。这时候，学生的行为使得谬误的谬误产生，即推动了元语言的病理循环。不管他在当下行动中掺杂了多少其它的价值，只要不是纯粹的追求真理的价值，似乎就可能产生病理性循环。值得注意的是，在这个例子中，学生并没有不认同规则在实践中体现的价值。价值并不一定是元语言产生病理性循环的根本原因。但是，学生在将谬误“武器化”的时候，确实产生了对元语言的误用，即谬误的谬误。这时，谬误的谬误的产生是源于学生在这个行动中的动机。因此，（b）应该改写为：

（b1）论证者在实践中的动机不是来源于积极价值（动机性问题）。

修改后的（b1）就能够解释学生的行为产生谬误的谬误的原因。因为他在这个行动中，动机并不是用谬误来通过论证追求真理，而是用谬误来赢得论证，或展现他的聪明，或表现他对谬误的活学活用。

然而，价值的不纯粹既可能产生谬误的谬误，也可能不产生谬误的谬误。考虑另外一个例子：假如有一个学生要参加一场辩论，他并不认为辩论就是为了追求真理，他只希望能够在这场辩论中成为赢的一方，让对方产生挫败感。为此，他研究了辩论的规则，寻找可以赢的方法，其中就包括正确使用谬误列表。谬误列表对他而言的价值是能够帮他取胜并打击对手，而不是追求真理。反之，对于对手而言，谬误列表的价值依然可以是追求真理。此时，两人所认可的同一个谬误列表的价值并不相同，但依然可以在辩论中正确使用谬误列表。[2] 因此，积极价值的认同也不是元语言不产生病理性循环的根本原因。

更进一步讲，一个源于认同论证积极价值的动机也可能产生谬误的谬误。以

[2]感谢匿名评审专家建议引入此例子，这有助于更直观地说明动机并不是谬误的谬误产生的根本原因的问题。

稻草人谬误（straw man fallacy）为例。稻草人谬误指的是歪曲对手的论点，从而更容易击败它。这种歪曲可能是弱化、虚构、断章取义等等。泰丽丝和艾金（R. Talisse & S. F. Aikin，[20]）、艾金和卡西（S. F. Aikin & J. P. Casey，[6, 7]）识别并定义了新型的稻草人变体，将其范围从简单的歪曲个别论点扩展到包括软弱人（weak manning）、空心人（hollow manning）和铁人（iron manning）等形式。由于稻草人谬误论证涉及两个论证者的相互攻击，沃尔顿和马卡尼奥（D. Walton & F. Macagno，[21]）指出，稻草人谬误可以更恰当地被视为一种辩证操作、策略或手段。稻草人谬误的多样化使其用法变得复杂。而且，在不同情境下，判定稻草人谬误是被误用还是被恰当使用也存在解释的空间。

考虑稻草人谬误的一种变体——软弱人。软弱人是指在论证过程中选择论证者最弱的点进行攻击，击败对方之后暗示关于该论点的结论的失败。软弱人夸大了其攻击的较弱的论点的重要性。然而，这样一种谬误也能够被合理利用。参考艾金和卡西（[7]）的关于同性婚姻的论证的例子。例子如下：

> 布拉德："我最近在保守派媒体上看到不少反对同性婚姻的论点。"
> 安吉丽娜："我也是。我从 RedState.com 的一位博主那里听到了一个特别糟糕的说法：他辩称，如果同性恋被允许结婚，没有什么能阻止他娶他的盒子龟。"[3]
> 布拉德：你这是稻草人谬误，你的论证是无效的。　（[7]，第 434 页）

在这个例子中，安吉丽娜回应了布拉德提出的论点中较弱的一个，是稻草人谬误中的"软弱人"的表现。布拉德指出这是谬误，并试图让安吉丽娜回答他的主要论证观点。他们的动机都是出于认同论证的积极价值的。从安吉丽娜的角度来看，她这么说并不是不合理的。艾金和卡西指出，优先回应最弱的论点，有助于进一步讨论。安吉丽娜的这种论证行为可以称为"清理场地"，即在转向更好的论点之前，先指出并应对普遍存在的糟糕论证。通过先解决糟糕的论点，使得讨论有逐步改善的空间。（[7]，第 434 页）

从布拉德的角度来看，当他将谬误用在对话中时，是将元语言对象化。同时，在这个行为中动机是出于认同论证的积极价值的。但是，由于谬误作为一种批评性元语言，是对论证行为的否定。安吉丽娜在这个例子中的行为并非是不合理的行为，那么她的这种回应论点的方式就不能够被成为谬误。如果根据稻草人谬误的标准，根据安吉丽娜回应的语言作出判断，将其识别为谬误，就会犯了错误识别谬误的问题，即谬误的谬误。

与此同时，与稻草人相对应的铁人也可能因为被误用。作为稻草人谬误的反面，铁人是指对论点的严肃辩护，不弱化、虚构或歪曲对方的观点。铁人的核心

[3]在艾金和卡西的文中，这个例子最后布拉德的回应是"哇，太搞笑了。"（[7]，第 434 页）本文将这个例子最后布拉德的回应改为"你这是稻草人谬误，你的论证是无效的"，通过新的回应来突出可能存在的问题。

在于用最大的善意来解释他人的交流行为。铁人不是谬误。然而，如果一个论证交流的过程中，有人试图用含糊不清的表达来引导对方对其不重要的观点进行善意的解释，这时候善意的解释极有可能被利用。采取铁人策略的人可能会耗费大量的时间和精力来对模糊的观点进行解答，而脱离问题的本身。艾金和卡西（[7]）举了一个课堂的例子，在例子中，教授在讲授笛卡尔第一沉思录的论证时提到了知识可能只是一场梦或者一个幻觉的观点。这时候一名学生就此联想到了自己因为喝太多止咳药配啤酒之后做的梦，并怀疑笛卡尔或许是跟他一样。学生的这个说法完全偏离了教授授课的主题。显然，在有限的课堂时间上花费时间去分析这个学生的观点是不合理的，最好的方式是不去深究，或者忽略这个观点。在这个情境中，如果根据铁人的论证要求，教授就应该花费时间对学生的观点进行分析和解释，但这些投入其实并不明智。艾金和卡西指出，这样“偏离主题的讨论不仅浪费时间，还会误导教育。”（[7]，第 437 页）那么，铁人在这个论证中就成了谬误吗？如果这时候有另外一个布拉德指出为不解释学生的这些观点是谬误，那么是否也犯了谬误的谬误呢？

从以上两个关于稻草人谬误的例子可以看到，论证者即便出于认同论证的积极价值的动机去使用元语言，他们也依然可能制造出新的谬误的谬误，从而衍生出关于谬误的病理性循环。那么，（b1）也是错误的，消极的动机和价值都不是产生病理性循环的必要条件。或许戈登会反驳这两个例子，认为“由我们的元语言词汇所引发的误解和错误，是可以通过熟悉的话语引导和修正技巧得到改善的。”（[14]，第 36 页）考虑到前面提出的问题是在不同论证情境中对谬误的使用，那么对谬误理论的使用限定不同的情境似乎可以解决这个问题。[4] 然而，情境具有特殊性和复杂性。对每个情境进行限定将使得论证规则将因此变得繁琐，从而失去规则本身的简洁性或普遍性特质。谬误使用者需要牢记大量的使用要求，并在每个情境中将各个条件一一对应，再加以使用，这使得能够使用这些元语言的人需要具备极高的知识素养。同时，对已发生的情境的总结并不能够涵盖未发生的情境。在未曾出现过的情境出现的时候，规范性从何而来就会成为新的问题。

由此可见，价值问题和动机问题都不是产生病理性循环的必要条件。即便避免了这两个问题，也可能产生病理性循环。那么，产生病理性循环的根源是元语言的结构性问题吗？不是的。问题在于论证的规范性。如同戈登所说，结构性问题可以通过“熟悉的话语引导和修正技巧得到改善”，但他忽略了一个问题，即什么才是“改善”，什么是好的论证，什么是坏的论证。如果元语言不能够对论证的好坏作出一个合理的评价，那么对元语言误用的病理循环将始终难以被消除。正

[4]支持语用辩证法的学者可能会认为规则在一定程度上是外在于特定受众或情境的，但并不是说规则与受众或情境完全无关，而是说这些规则及其应用在受众或情境变化时并不会发生太大变化。本文并没有完全否定语用辩证法的观点，仅仅指出规则在实践情境中可能遇到的困难。

如上一段所解释的，在不同的情境中，同样的论证方式、论证语言可能是好的，也可能是坏的。而评价一个论证的好坏，则需要考虑论证的规范性。尽管艾金和戈登在分析谬误时也考虑了语境和对话结构，但他们对规范性的最终溯源很大程度上仍停留在“元语言”的规则制定层面。换言之，他们试图通过完善语言互动的规则（如批判性讨论规则）来规范论证。然而，本文认为，论证不仅仅是遵循语言规则的对话游戏，更是一种植根于人类认知本能的适应性活动。仅依靠元语言规则的约束，往往难以解释为何某些谬误（如谬误的谬误）在认知层面具有顽固的诱惑力。除了显性的语言规则外，那些潜藏在认知机制、社会信誉评估及进化适应性中的“自然规范性”因素，虽然不直接体现为语言规则，却对论证的成败起着更为基础的制约作用。

4　论证的自然规范性

什么论证规则在什么情境下是适用的？卡斯特罗（D. Castro，[9]）预设了一种适用于论证规则的标准情境，将日常论证的、区别于标准情境的情境称为次优情境。他认为通过修复和补偿，能够使得次优情境接近标准情境，并在之后继续使用批判性论证规则来讨论问题。然而，他无法说明什么才是足够接近标准情境的次优情境。因此设定标准情境的想法很难实现。因此，这不仅仅是为每套论证规则建立一个情境的设定作为辅助规则的问题，而是需要先回到为什么这套规则在这样的情境中适用的问题。这就是论证规则的规范性来源问题。在论证理论中，不同的立场持有者对论证的规范性来源有不同的定义，而每种定义都有其局限性。[5] 在上述讨论中关注谬误的是非形式逻辑（Informal Logic）的立场。非形式逻辑诉诸相应语境下语言上的谬误，认为对谬误的规避就是对规则的遵守。规范性来源于是否规避了谬误。根据上述的分析，这种规范性是存在问题的。一方面，在确定规范性之前需要对谬误是什么作出判断，而判断又需要回溯到“为什么这是谬误”的规范性判断中，从而形成无限的循环。另一方面，它强调论证中的语言部分，狭义地将情境预设为外部条件约束论证者。

迈克尔•吉尔伯特（M. A. Gilbert）在他 2002 年的文章（[12]）中提出逻辑中心主义谬误（logocentric fallacy），认为将语言的逻辑化的表现形式视为唯一合理的交流方式是一种谬误。他认为，语言本身并不一定具有清晰的意义，语言所传达的信息对于使用它并且熟悉相关问题和背景的人来说是清楚的，但对不熟悉的人而言则是模糊的。论证交流作为一种信息传递的活动，如果过度关注语言方面，会限制获取和传递信息的实际内容。

在上文分析的艾金和戈登的观点中，他们认为谬误是用来评价论证好坏的元

[5]这里不展开讨论。参见吉尔伯特在 [13] 的批评。

语言。论证者在学习了谬误的用法之后，可以利用它去实现自己的论证目的，并产生谬误的谬误，这些行为都是在论证语言的抽象层面上进行的。将论证中的语言提取出来，仅考虑语境，忽略参与者更广义的背景、关系、目的等因素，直接对论证的好坏作出判断就产生了问题。假如为学习了谬误理论而产生谬误的谬误的学生赋予背景和目的，例如：

背景一：他是一个学习成绩很差的学生，总是掌握不好各种知识点。

背景二：他所论证的对手是一个聪明的学生，并且经常因为自己的聪明而霸凌其他同学。

背景三：他的论证目的是为了帮助某个被霸凌的同学，打击这个聪明的学生。

这些背景内容并不会通过论证语言体现出来。根据这个背景内容的添加，产生谬误的谬误并不会被认为是一个失败的论证。从整体上看，产生谬误的谬误体现了这个学生正义和机智的一面，他可能会因此受到赞扬。谬误在这里并没有起到一个引导或者指导论证如何进行的作用，而这恰是论证规范性所必备的。因此，如果以语言上的谬误为论证规范性的来源，会造成对论证活动中其它因素的忽略，导致了论证规则在不同情境中的不适用的问题，从而产生“对规则的误用”的病理性循环。由此可见，基于谬误的规范性来源的局限性在于对论证中其它因素的忽略。

那么，应该考虑哪些因素呢？吉尔伯特提出的关于论证的自然规范性[6]（Natural Normativity）可以回答这个问题。吉尔伯特（[13]）提出一个包含目标（goal）、情境[7]（context / situation）和信誉（ethos）三个部分的动态框架作为论证活动的自然规范性。他指出：

> 这三者共同构成一个对论证进行规范控制的系统，这个系统比任何抽象规则都更强大、更持久。当然，基本事实是，论证就像生活一样，是一种社会行为。作为一种社会活动，我们受到无数力量的引导和控制，而我们平常往往未曾意识到这些力量。这些力量源于我们的目标、语境、对自身信誉的感知，以及论证伙伴的信誉的复杂交织。 （[13]，第 7 页）

他强调了论证活动所受到影响的因素的复杂性，并指出论证规范应该是具体的，是在每一个实际的论证活动中产生的，而不是抽象的或绝对的。这三者之间没有明确的界限，但共同作用于论证活动，成为产生规范性的自然来源。

论证中的目标是多重的。吉尔伯特认为从表面上看，论证的目标就是说服对方。然而，除此之外，往往包含着其它目标，如面子目标（face goals）和关系目标（relationship goals）。（[13]，第 3 页）面子目标是指利用论证活动来维护自己的形象。关系目标是指通过论证活动来维护某种关系。这些不同的目标构成一个复合

[6]此处的“自然”是指自然而然地源自于目标、语境和信誉三大核心要素。（[13]，第 8 页）

[7]吉尔伯特在文中说明 context 和 situation 在他的观点中是可以互用的，本文统一翻译为情境。

的目标体系，会引导论证者的论证活动，限制他们在论证中的策略和选择。在戈登的论证中，他提到当论证者将谬误作为巩固自己的认知偏差的工具时，就可能出现谬误的谬误。这就是由目标在论证中对行为者的引导。对谬误的不同使用方法，是论证者被自身的论证目标所引导而自然产生行为。要判断这个论证行为是否符合论证的规范性，目标是需要被考虑的部分。但是，目标所指导的规范性只是一部分，还需要考虑另外两个因素的作用。

影响论证规范性的另一个因素是情境。情境强调论证应该以具体的现实为基础。吉尔伯特认为的论证情境包含但不限于参与者之间的关系、互动地点，以及在论证中起到作用的政治、社会和经济等背景因素。同样一个论证发生在不同的人之间，或者在不同的地点、不同的背景下，所产生效果或者意义都显然不同。在某一领域中被归为谬误的表现，可能在另一领域中并非如此。就像在上述第三部分中讨论到的例子，铁人在课堂情境下可能会变成谬误，而软弱人谬误在某些特定的情境下会产生正向效果，这都是源于情境对论证规范性的影响。另外，参与者之间的关系是情境因素的核心。参与者之间共享哪些信息、彼此的熟悉程度都是影响情境的重要部分。对于彼此熟悉且共享信息较多的参与者来说，他们之间的论证可以不展开解释很多内容，甚至一个眼神或者手势就能够传递信息。反之，较为陌生的参与者在论证时则需要对很多的内容作出解释。

根据吉尔伯特的观点，情境是一个较为广义且模糊的概念，难以明确界定哪些内容应该被包含在论证情境里。吉尔伯特认为这是具有好处的：模糊性的另一面是“赋予了分析者观察实际发生的事情的灵活性，而不是拘泥于严格的指南或框架”。（[13]，第 5 页）吉尔伯特并没有给出情境如何产生规范性的方案，但指出情境是影响规范性的因素。他认为论证的规则是可以变化的，这种变化是随着情境的改变而改变的。不仅如此，情境和目标之间是互相影响的。

第三个影响论证规范性的因素是信誉。信誉指的是互动参与者的可信度和可靠度。信誉在论证的规范性中不仅仅指公众人物或专家权威的可信度，还指普通的人与人之间的信任。论证的参与者对彼此的信誉评价会随着交流的增加而发生改变。人们很难在交流中对信誉持一个完全中立的态度。信誉一旦发生改变，参与者之间的关系就会发生改变，或者更确切地说是论证的情境会因此改变。例如，当一个人发现别人多次出道听途说，那么在与其进行论证交流时，便会对对方的观点做更仔细的考察。如果经过几次交流发现对方提出的每个想法都是经过严谨考察才会说明的，那么在论证交流中可能会更容易相信对方的论据。同时，他人的信誉判断也会影响参与者对自身声誉的认知。参与者的行为会受到希望被他人以好的信誉来评价的愿望所驱使，从而形成在论证中维护声誉的目标。之前所提到的人的互动性，在信誉因素中得到了体现。

考虑人际关系作为论证规范性基础，除了吉尔伯特所提出的信誉，在其他学

者那里也得到了重视。斯洛布（W. H. Slob）指出，论证的规范性应该考虑对话的可理解性、可接受性和可回应性。论证能够在两个论证者之间进行下去是论证的规范性的重要考虑因素。

> 对话式修辞真正具有论证性，仅涉及参与者各自的实质性贡献。任何结论的规范力，都是讨论过程中的举动所带来的结果。
>
> （[19]，第 191 页）

杰克逊（S. Jackson）认为自然规范性应该体现在人们共同管理分歧时对彼此的问责和协作。

> 自然发生的论证以分歧管理为核心组织，完全取决于参与者在彼此的行为上遵循并执行标准的能力。最根本的是，参与者期望彼此在分歧的管理上进行合作，例如在预期将产生实际后果时表达异议，并以有助于其管理的方式回应已表达的异议。（[16]，第 642 页）

通过目标、情境和信誉三个因素，吉尔伯特的自然规范性观点考虑了论证活动中的几乎所有因素。大致可以梳理为：目标体现了论证的内容的可能性，情境体现了论证的背景，而信誉则是考虑了论证参与者之间的关系。这三个部分的内容之间并不能划出清晰的界限，而是互相影响互相交织，同时也互相制约。这样的规范性是全面的、动态的、考虑情境的。这种自然规范性避免了仅仅依靠语言规则而机械行事所产生的困难。是情境决定了规则的选择，而不是试图用规则去规范情境。以自然规范性作为规范的来源，充分考虑这样的规范，就避免了在不同情境中规则无法被掌握而产生滥用、误用的可能。

语用辩证法中，也同样存在考虑情境的论证概念——战略操控（Strategic Maneuvering）。[8]爱默伦和胡特洛瑟（F. H. van Eemeren & P. Houtlosser，[11]）提出战略操控的概念，认为论辩者在试图遵守批判性讨论规则以解决分歧的同时，也会通过选择最优的话题潜力（topical potential）、迎合受众需求（audience demand）以及利用表达手段（presentational devices）来追求己方立场被接受的最大化。战略操控的观点在考虑论证的规范性时，充分考虑了情境的重要性。然而，这与吉尔伯特的自然规范性观点存在较大的区别。

两种观点中，情境对规范性的产生的影响不同。战略操控的规范性源于外部预设的规则，这些规则就像足球比赛的章程，参与者必须遵守才能“玩这个游戏”。与之不同，吉尔伯特是将情境是为一种“基于现实的建构”（Reality-grounded construct），不仅包含规则、制度背景，还涵盖了参与者之间的具体人际关系、地理位置以及动态变化的社会政治因素。在战略操控中，情境起到了选择规则的作用；

[8]感谢匿名评审专家提示笔者关注战略操控观点与自然规范性观点的异同，本部分的讨论受益于该建议。

在自然规范性下，情境参与了规范性的构成。战略操控是一种抽象规则，只是在规则的使用上考虑了情境因素。反之，自然规范性是由“目标、情境和信誉”构成的复合体，在人际论证中由这个复合体自然产生有效的制约和导向，是具体的和动态的。

以歌手参加比赛为例。假设有个歌手去参加比赛，但他的目的是提高自己的知名度，而不是赢得比赛。从战略操控的角度来看，歌手的“规范性”来自于比赛的制度先决条件（Institutional Preconditions），他必须遵守规则的同时，尽可能表现得优异以打动评委。歌手会从曲库中选择最能发挥自己优势的歌（选题潜力），根据评委的喜好调整演唱风格（受众需求），并运用华丽的舞台包装（表达手段）。根据战略操控的分析，他只能够在这个具体情境的规则边界内追求成功，当他因为一些意外（例如唱跑调、受伤等）而知名度得到提高，在这战略操控的角度看就会被视为“出轨（derailment）”或谬误。

相比之下，自然规范性的观点将歌手的目的（即提高知名度）纳入他的行动规范性中。在吉尔伯特的框架下，目标是多元的，而不仅仅是“赢得比赛”。除了“赢得比赛”这个战略目标，歌手可能还有关系目标（例如结识评委）或面子目标（保持自己高冷的艺术形象）。同时，歌手不仅仅是遵循比赛规则，更是在维护自己作为一名“专业音乐人”或“独立艺术家”的信誉。这种信誉感会像一种“自然的压力”一样限制他的行为，使其不至于为了赢而不择手段。唱歌比赛只是情境的一部分，规范性来源也不是纯粹的比赛规程，而是由歌手与其他参赛者、评委、观众之间的互动关系构成的动态现实。简单而言，战略操控就像是歌手在读一本《参赛指南》，他在研究如何在不扣分的前提下拿到最高分；而自然规范性则是歌手在经营自己的职业生涯，他不仅看重当下的分数，更看重这一场表演在他整个人生目标和社交信誉中的价值。因此，根据吉尔伯特的观点，歌手短期的失误并不会被解释为“出轨”或谬误，而可以是构成自然规范性的目标的一部分。

回到本节最初提到的问题：什么论证规则在什么情境下是适用的？这并不是一个为每一种类似的情境适配一个论证规则的问题，而是考虑的是情境如何成为、构成论证的规范性来源的问题。规范性是“自然产生的”，其限制来自于行动者对声誉的保护以及对他人作为自主个体的尊重，而非抽象的逻辑规则。

5　从自然规范性看密涅瓦猫头鹰问题的局限性

当关于论证的规范性来源被厘清，关于密涅瓦猫头鹰问题的局限性就同时得以呈现。一方面，由于论证的自然规范性是动态、多元的，是根据不同的目的、语境和信誉产生的，对于论证规则的使用是否合理的元语言就具备了灵活性。失去了抽象的普遍标准，元语言因为没有恰当识别情境的误用的可能性得到了降低。以

前述稻草人谬误为例，在充分考虑论证中所涉及的目的、语境和信誉之后，安吉丽娜的论证并不会被认为是不好的，布拉德对谬误的指出也不会轻易被否定。同样，学生制造谬误的谬误的行为被充分考量之后，也不会直接定义为好的或者坏的，而是会根据他的实际目的、他所处的情境以及他的信誉来评估他的论证。

另一方面，由于信誉的约束，故意将元语言武器化用于攻击别人的代价被提高，这个问题产生的可能性也因此同样变小。区别于目的和情境，信誉是一个累积性、持久性的考虑因素。论证者通过一次又一次地与人交流来形成自己的信誉，参与者会根据对论证者的信任度来判断他所提出的论据的正确性，或者考虑是否与他继续论证。回到艾金（[5]）最初在提出病理循环问题的时候举的一个例子：足球比赛。由于足球比赛中，球员常常在靠近自己防守的球门时，用粗暴的方式抢球。为了改变这个情况，足球比赛设定了“如果进攻方在对方禁区内被犯规，球队就会获得一个点球”的规则。但因为有了这个规则，进攻球员会假装被犯规，即假摔，从而获得点球机会。艾金认为这是一个密涅瓦猫头鹰的问题。然而，当信誉在这个例子中被考虑。如果一个球队在比赛时使用了假摔，他们的信誉就会相应地降低。其他球队可能会对其有所防备，或者拒绝跟这样的队伍进行比赛。那么，球队在考虑是否假摔时就有了更大的压力。如果从这个角度考虑，故意对元语言进行误用的概率就会降低，病理性循环问题就能够得到限制。论证理论同理。

信誉的约束也体现了论证理论的另一个问题：论证评价应该是独立的还是连续的？传统意义上，好的论证是一个独立实践，这就使得论证者可以在一个论证中独立拥有一个目的，而不对他的其它论证行为产生影响。把每个论证孤立开来，单独评价它的好坏，确实可以将它们作为学习如何论证的某种案例。但是，在实践中，当论证者作出一个论证时，他所表现出来的论证得“好”可能是偶然性的。这甚至难以说明该论证者掌握了论证理论，他之后的论证不一定就能是“好”的。如果对论证的评价始终是孤立的，密涅瓦猫头鹰问题就会显现。当论证者在一个论证中只考虑当下论证能否赢，不需要考虑到评价对他今后的影响，那么他极有可能会不计后果地把元语言当作用来赢得论证的武器。而一旦对当下的论证的评价会影响到他未来的论证，特别是他的信誉时，他就不会轻易地作出这样的举动。产生病理循环的问题也会因此得到限制。

正如阿伯丁和科恩（[3]）在德性论证理论所讨论的，论证实践中，论证者的角色所带来的影响难以被抹掉，对论证的好坏的评价应该能够说明它对论证者产生了什么影响，并预测论证者未来的论证。因此，对于论证的评价应该考虑的是连续性实践，即通过当下的论证，论证者能否继续获得论证的机会，而不是以失去他人的信任或继续对话的机会为代价。一旦考虑到连续性评价的问题，信誉就会被重视，论证规范性自然会考虑到信誉的作用。根据以上的讨论，对德性论证理论与论证者品格好坏无关的批评也不攻自破。阿德勒（Jonathan Adler，[4]）批评认为，

德性论证者的品格的好坏，与评估论证强度毫无关系，因为这完全取决于前提与结论的关系。这就回到了对论证仅仅考虑语言逻辑和单独事件的问题讨论。阿德勒认为的这种评估方式就是产生论证的密涅瓦猫头鹰问题的根源所在。因此，将论证嵌入到更大的理性和认知生活的背景中，强调论证的道德和品格层面（[2]），这即是对信誉作为自然规范性的一部分的充分体现，同时也是避免论证元语言产生病理性循环的方法。

6　结语

本文从艾金提出的密涅瓦猫头鹰问题出发，通过戈登对该问题的分析，试图探索问题的根源。艾金认为由于元语言与对象语言之间虽然有语义距离，但是这两者在实践中经常会被混用，元语言可能被当作对象语言使用，从而产生新的关于元语言的元语言。这个问题在论证中体现为产生了谬误的谬误、谬误的谬误谬误……的病理性循环。戈登通过分析艾金提出的论证的元语言问题，认为这并不是元语言的结构性问题，而是论证者对论证规则的价值认同问题。

本文重构了戈登的论证，认为价值认同并不是问题的根源，动机也不是产生这个问题的必要条件。本文指出，对论证的规范性的狭隘的考量才是产生病理性循环问题的根源所载。随后，本文引入吉尔伯特的自然规范性观点，通过从目的、情境和信誉三个因素的结合所产生的自然规范性来理解对论证的评价。在这个基础上，发现了自然规范性对密涅瓦猫头鹰问题的限制。一方面是避免了普遍标准对实践复杂情境的不适性，另一方面是通过信誉约束论证者的行为。最后，本文延伸了对关于论证的密涅瓦猫头鹰问题的思考，认为这个问题与强调独立评价论证有关。要避免这个问题还应该注意论证评价的连续性问题。

参考文献

[1] A. Aberdein, 2023, “The fallacy fallacy: From the owl of minerva to the lark of arete”, *Argumentation*, **37(2)**: 269–280.

[2] A. Aberdein and D. H. Cohen, 2016, “Introduction: Virtues and arguments”, *Topoi*, **35(2)**: 339–343.

[3] A. Aberdein and D. H. Cohen, 2024, “Virtue theories of argument”, *Inquiry: Critical Thinking Across the Disciplines*, **33(2)**: 117–142.

[4] J. E. Adler, 2007, “Commentary on daniel H. Cohen: ‘Virtue epistemology and argumentation theory’”, in H. V. Hansen, C. W. Tindale, J. A. Blair, R. H. Johnson and D. M. Godden(eds.), *Dissensus and the Search for Common Ground*, pp. 1–5, OSSA.

[5] S. F. Aikin, 2020, “The owl of minerva problem”, *Southwest Philosophy Review*, **36(1)**: 13–22.

[6] S. F. Aikin and J. Casey, 2011, “Straw men, weak men, and hollow men”, *Argumentation*, **25(1)**: 87–105.

[7] S. F. Aikin and J. P. Casey, 2015, “Straw men, iron men, and argumentative virtue”, *Topoi*, **35(2)**: 431–440.

[8] S. F. Aikin and R. B. Talisse, 2019, *Why We Argue* (*And How We Should*), New York: Routledge.

[9] D. Castro, 2022, “Argumentation in suboptimal settings”, *Argumentation*, **36(3)**: 393–414.

[10] C. Cotton, 2018, “Argument from fallacy”, *Bad Arguments: 100 of the Most Important Fallacies in Western Philosophy*, pp. 125–127, Chichester: John Wiley & Sons.

[11] F. H. van Eemeren and P. Houtlosser, 2002, “Strategic maneuvering in argumentative discourse: Maintaining a delicate balance”, in F. H. van Eemeren and P. Houtlosser(eds.), *Dialectic and Rhetoric: The Warp and Woof of Argumentation Analysis*, pp. 131–159, Dordrecht: Kluwer.

[12] M. A. Gilbert, 2002, “Effing the ineffable: The logocentric fallacy in argumentation”, *Argumentation*, **16(1)**: 21–32.

[13] M. A. Gilbert, 2007, “Natural normativity: Argumentation theory as an engaged discipline”, *Informal Logic*, **27(2)**: 149–149.

[14] D. Godden, 2022, “Getting out in front of the owl of minerva problem”, *Argumentation*, **36(3)**: 35–60.

[15] I. Hacking, 1999, *The Social Construction of What?*, Cambridge: Harvard University Press.

[16] S. Jackson, 2019, “Reason-giving and the natural normativity of argumentation”, *Topoi*, **38(4)**: 631–643.

[17] W. G. Lycan, 1997, *Consciousness and Experience*, Cambridge, Mass.: MIT Press.

[18] C. K. Miller and J. S. Miller, 2015, *Why Brilliant People Believe Nonsense: A Practical Text for Critical and Creative Thinking*, Acworth, GA: Wisdom Creek Academic.

[19] W. H. Slob, 2002, “How to distinguish good and bad arguments: Dialogico-rhetorical normativity”, *Argumentation*, **16(2)**: 179–196.

[20] R. Talisse and S. F. Aikin, 2008, “Two forms of the straw man”, *Argumentation*, **20(3)**: 345–352.

[21] D. Walton and F. Macagno, 2010, “Wrenching from context: The manipulation of commitments”, *Argumentation*, **24(3)**: 283–317.

（责任编辑：执子）

The Owl of Minerva Problem: Dilemma and a Natural Normativity-Based Resolution

Lingxin Cai

Abstract

Fallacies are a crucial component in the study of argumentation within informal logic. Yet, as Aikin raises in his lecture, a concern arises: once fallacy as a meta-language becomes possessed by agents with interactive dispositions—humans—the fallacy fallacy can be triggered, which may further generate a fallacy of fallacies, hence giving rise to a pessimistic Owl of Minerva Problem. Godden deconstructs this issue and contends that the cause is not a structural problem but a motivational one; from an optimistic vantage, this suggests auxiliary rule-correction to avoid much misapplication. This paper argues that motivation is not the cause of the Owl of Minerva Problem; rather, the essence lies in neglecting information beyond linguistic form in arguments and failing to give sufficient consideration to the normative sources of the argumentation rules. By drawing on Gilbert's natural normativity theory, this paper investigates the misuse of fallacies and, in light of that framework, examines the limitations and possible solutions to the Owl of Minerva Problem.

Lingxin Cai　Department of Philosophy, Sun Yat-sen University
lx_cai@163.com