悖论的生成机制和解释

吴美华 王拥军 杨义川 王潇扬

（北京航空航天大学数学与系统科学学院 北京 100191）

摘 要 从计算机科学中的具体悖论实例出发，使用对角线方法来说明一类悖论的生成机制，并指出自指代现象是悖论产生的深层次原因。传统的应对策略往往采用回避的方式，简单禁止自指代以避免悖论。从量子力学和范畴理论两个新视角出发，给出容纳悖论的新模型。结果表明，从新角度审视悖论不仅可以使悖论在某些新领域得到合理解释，而且能提供认识问题本质的新思维。

关键词 悖论，自指代，对角线方法，量子力学，范畴理论

中图法分类号 TP301，O141 **文献标识码** A

Generation Mechanism and Interpretations of Paradoxes

WU Mei-hua WANG Yong-jun YANG Yi-chuan WANG Xiao-yang

（School of Mathematics and System Sciences，Beihang University，Beijing 100191，China）

Abstract Based on some concrete paradoxes in computer science，this paper used the diagonal arguments to explain the generation mechanism of a class of paradoxes，and pointed out that a deep reason of resulting in some paradoxes is the self-reference. Moreover，from two new perspectives in quantum mechanics and category theory，this paper gave some interpretations including paradoxes rather than traditional method to avoid a paradox by prohibiting the self-reference. It shows that our models not only can provide a reasonable explanation for some paradoxes，but also can provide some new ideas to understand the essence of the paradoxes.

Keywords Paradox，Self-reference，Diagonal arguments，Quantum mechanics，Category theory

1 引言

悖论指的是在逻辑上可以推导出互相矛盾的结论，但表面上又能自圆其说的命题或理论体系。在数学中，悖论的抽象公式就是 $A \Leftrightarrow \neg A$。悖论看似自相矛盾，其实往往揭示了某些重要问题的真实本质。

悖论是科学研究中不可回避的问题。康托于 19 世纪末创立的朴素集论逐渐成为了数学大厦的基础。随着人们对集合论的深入了解，集论本身也暴露出了一些问题，最著名的就是罗素于 1902 年提出的罗素悖论，这从根本上撼动了朴素集论的根基地位，引发了数学史上的第三次危机[1]。人们意识到集合论并不完善，并在之后提出了公理集论系统。理论数学中的塔斯基定理、哥德尔不完全定理以及理论计算机科学中的停机问题[2]，都与悖论有着密切的联系。

探究悖论能让我们更加清晰地认识事物的本质，发现新的问题研究领域。国内外学者对悖论的研究[3]主要有以下几个方面：罗素的简单类型论和分支类型论、塔斯基的语义层次理论、鲍契瓦的三值逻辑、克里普克的真值间隙论[4]等，但这些都不能圆满地解决悖论问题。近年来，众多学者在不同的领域研究悖论问题，极大地深化了对悖论的认识。范畴论[5]是研究对象之间保持结构关系的数学，研究一般化的所有对象，是抽象刻画数学对象的有力工具。Yanofsky[6]于 2003 年在卡积闭范畴中利用不动点定理解释了多个自指代悖论；量子力学[7]也是解释悖论的一种新的研究工具，在量子力学中存在波函数，其将量子模型与人类认知理论联系起来[8]。关于悖论和自指代的关系，Yablo[9]于 1993 年提出了基于无限关系的非自指代悖论；Priest[10]于 1994 年指出了一类自指代悖论的统一结构；SMITH[11]于 2000 年提出了解决自指代悖论的统一方法；Parent[12]于 2016 年说明语义悖论均由自指代所导致。

本文第 2 节从计算机科学中一些具体的悖论实例出发，使用对角线方法来说明涉及自指代否定的问题必然导致悖论；第 3 节给出容纳悖论的模型，对悖论进行新的解释；最后总结全文。

2 一类悖论的生成机制

本节以集合论中的罗素悖论、离散数学中的理发师悖论、说谎者悖论以及计算理论中的停机问题为例，使用对角线方法来探究这类自指代悖论的生成机制。

对角线方法[9]可描述为：设 R 是集合 A 上的二元关系，$D=\{a\in A\mid (a,a)\notin R\}$ 称作 R 的对角线集合，对于每一个 $a\in A$，令 $R_a=\{b\in A\mid (a,b)\in R\}$，则 D 与每一个 R_a 都不相同。简单来说，对角线上的补元素一定与每一行对应元素不同。

作为一种证明方法，对角线方法构造出了一个超越可数无限的“新对象”。因此，可以说，对角线方法实现的是质的变化，从现有的层次到达一个更高的层次，是从系统之外看待问题的一种方式。康托定理证明形象地体现出了这一点。首

本文受国家自然科学基金（11271040），北航凡舟教学团队建设资助。

吴美华（1993－），女，硕士生，主要研究方向为理论计算机科学、机器学习，E-mail：2271586170@qq.com；**王拥军**（1970－），男，博士，讲师，主要研究方向为软计算、范畴理论、数据挖掘；**杨义川**（1970－），男，博士，教授，主要研究方向为软计算、逻辑代数；**王潇扬**（1994－），男，硕士生，主要研究方向为为软计算、机器学习，E-mail：420478596@buaa.edu.cn（通信作者）。

先,假设从自然数集到自然数集上的函数全体是可枚举的,得到一个函数可枚举列 $f_0,f_1,f_2,\cdots$。通过改变对角线上的元素构造出函数 g,$g(n)=f_n(n)+1$,不妨设 g 在枚举列中为 f_k,此时自指代作用在其编号 k 上,会得到 $g(k)=f_k(k)=f_k(k)+1$,矛盾。因此,g 作为"新对象",在形式系统内部是不能处理的,否则会出现不一致。

对角线方法在计算机科学中有着广泛的应用,著名的塔斯基定理和哥德尔不完全定理的证明都两次用到了对角化思想。以下用计算机科学中的实例来说明其在自指代悖论中的关键作用。

2.1 罗素悖论

1902年,罗素构造了一个"集合"S,$S=\{a|a\notin a\}$。考虑 S 是否属于它自身时,会得到 $S\in S\Leftrightarrow S\notin S$,导致悖论。

对角线方法可以形式化地揭示导致该悖论的原因。由 S 的定义可知,任意集合 a 属于 S 当且仅当 a 不属于自身,即 $a\in S\Leftrightarrow a\notin a$(第一次对角化),通过"所有不属于自身"的集合性质构造出一个新"集合"S。当取集合 a 为 S,且 S 作用到自身时,得到 S 属于 S 当且仅当 S 不属于自身,即 $S\in S\Leftrightarrow S\notin S$(第二次对角化),导致悖论。可以看出,集合属于关系的两次对角化导致罗素悖论的出现。解决此悖论的传统公理集论认为,这样构造的 S 不是集合,而是真类。

罗素悖论的产生机制同样适用于理发师悖论。

2.2 理发师悖论

在一个小镇上,有一个理发师,他只给镇上不给自己理发的人理发,问题是他该不该给自己理发?

设理发师 A 给某个人 X 理发,表示为 AX。而 A 只给镇上不给自己理发的人理发,则表示为 $AX\Leftrightarrow\neg XX$(第一次对角化)。当理发师 A 的理发行为作用到自身时,就表示为$AA\Leftrightarrow\neg AA$(第二次对角化),导致悖论。归结于理发行为的两次对角化必然导致理发师悖论。为了避免理发师悖论,解决此悖论的传统方法认为,这样的小镇不存在,没有这样的理发师。

2.3 说谎者悖论

说谎者悖论是另一个著名的语义悖论,是典型的由于语言层次不分明导致的自指代悖论。有多种形式,比如"我正在说谎""(1):(1)是错的"等。

以"我正在说谎"为例,假设"我"在说谎,则该语句内容"我正在说谎"是个谎言,又推出"我没说谎";如果"我"没说谎,则语句内容为真,即"我正在说谎",导致悖论出现。通俗地讲,内容里的"说谎"是一种否定,即第一次对角化;命题的内容涉及对自身的判断,即第二次对角化。语句真值的两次对角化导致说谎者悖论的出现。为了避免出现类似的悖论,传统的解决方法认为"我在说谎"这样的表述无意义。

2.4 停机问题

停机问题:是否存在这样一个停机程序 $halts(P,X)$ 能正确地判定在输入 X 上程序 P 是否停机?

图灵停机问题一直是理论计算机科学中重要的问题,使用两次对角化过程可以给出解答。

第一次对角化,用它编写一段程序 $diagonal(X)$:

a:if $halts(X,X)$ then goto a else $halt$

现在第二次对角化,问题是 $diagonal(diagonal(X))$ 是否停机?答案是奇怪的。它停机当且仅当 $halts(diagonal(X),diagonal(X))$ 调用为否,也就是不停机,导致悖论。则说停机程序不存在。

事实上,对角化方法可以简化为两步:第一次对角化 $AX\Leftrightarrow\neg XX$;第二次对角化是取 A 为 X,得到 $AA\Leftrightarrow\neg AA$。两次对角化过程必然导致悖论。

2.5 悖论的生成机制

所有上述这些例子都说明了一个问题:当涉及自身判断时,就会出现问题。任何事物都不能"讨论"自己的真值或"描述"自身性质,也就是自指代思想。自指代现象才是这类悖论产生的深层次原因。

然而在系统中通过禁止自指代来避免悖论的出现,是一种消极的态度。在罗素悖论中,把一部分"集合"解释为真类(proper class)以保证集合论的一致性;解决理发师悖论时,就简单地认为不存在这样的小镇;对于说谎者悖论,这种自指代语义悖论被看作是无意义的表述;在停机问题中,认为停机程序不存在。用这类限制的方法避免悖论的出现,并没有从正面接受悖论确实存在。以下将从新的视角,尝试找到容纳悖论的新模型。

悖论并不是矛盾,随着我们认识的深入,其在塔斯基定理和哥德尔不完全定理中发挥着积极的作用。塔斯基定理,类似于说谎者悖论,可看作是说谎者悖论在形式语言下的不可翻译。塔斯基定理,说明在任何一致的算术系统中不能定义自身的真值谓词。而在哥德尔不完全定理的证明中,"该命题不可证明"的论述可看作是将"真"变为"可证"的说谎者悖论在机器语言下翻译的结果。哥德尔不完全定理,说明任何包含算术的数学系统都不能完备地讨论他们自身的可证性。悖论进一步深化了我们对问题的本质认识。

3 悖论的解释

悖论的出现,是因为自指代导致了真假值不相容的局面。因此,我们试图找到新的领域,以更好地处理自指代问题,解释真假共存现象,从而找到容纳悖论的模型。在量子力学中,将悖论看作是一种认知实体,测量之前处于一个亦真亦假的叠加态中。当我们对其认知测量时,才会坍缩成某一本征态,真或假。同样地,在范畴理论中,将真假与特定时刻对应起来,某一语句的真值并不是固定不变的,而是随时间演化的,这一时刻为真,下一时刻可能为假。利用范畴理论中的集合值函子(时序集合)建立容纳悖论的新模型。基于新的视角,悖论也提供给我们认识事物的新思维。

3.1 量子力学视角

量子力学中的核心概念——波函数为解释悖论提供了新的工具。科学家利用波函数来计算粒子拥有某种特性(如处于某个确定的位置而非其他位置)的概率。波函数并非是真实存在的,正是测量本身使得被测量的特性成为真实存在。

量子理论认为,对一个量子的测量会对它本身的状态产生干扰,测量得到的结果不是量子本身的状态,而是干扰后的状态,会坍缩成量子的本征态。类似地,在说谎者悖论中,如果把说谎者悖论看成一个认知实体,每次读句子和进行假设的过程就是一个测量的过程。

对于单个说谎者悖论"(1):(1)是错的",可以从量子力学的角度进行解释。

以量子力学中自旋1/2的粒子为例,该粒子具有两个本征态:自旋向上和自旋向下。它在恒定磁场中的振荡与说谎者悖论命题真假的判定有异曲同工之妙。因此,根据自旋1/2的粒子表示方法,在希尔伯特空间里,Ψ 表示说谎者悖论原本的状态:

$$\psi=c_{\text{true}}\begin{pmatrix}1\\0\end{pmatrix}+c_{\text{false}}\begin{pmatrix}0\\1\end{pmatrix}$$

对应的投射算子 P_{true} 和 P_{false} 分别为：

$$P_{\text{true}}=\begin{pmatrix}1&0\\0&0\end{pmatrix},P_{\text{false}}=\begin{pmatrix}0&0\\0&1\end{pmatrix}$$

在量子力学描述下，在叠加态 Ψ 下进行真值测量后，改变了原来的状态，结果为真或假时，表述如下：

$$P_{\text{true}}\psi=c_{\text{ture}}\begin{pmatrix}1\\0\end{pmatrix},P_{\text{false}}\psi=c_{\text{false}}\begin{pmatrix}0\\1\end{pmatrix}$$

因此，从认知实体的角度，认为说谎者悖论作为一种认知实体，与量子具有相似的上下文相关性质，即测量的过程会对测量对象产生干扰，利用自旋 1/2 粒子的自旋向上和向下分别描述一个说谎者悖论的真假情形，避开了矛盾对立面真假不相容的局面。

在量子力学的视角下，利用量子的叠加态，说谎者悖论真假值得以共存，避免了不相容的局面，得到了容纳悖论的一种模型。

3.2　范畴理论视角

范畴论专注于对象间相关的结构关系——态射，即“箭头理论”，而不关注对象内部的结构。函子[14]是范畴间的态射，由于结构保持性，集合值函子可以将复杂的范畴通过态射作用到集合范畴，再在集合范畴里解释其结构性质。一个句子的真假在某个特定的时刻是确定的，将某个时刻与特定的阶段对应起来。在不同的阶段，语句得到的真值可能不一致。

具体来说，在集合范畴中，用映射来表示集合元素更具有一般性。集合中的元素用单点集{ * }(终对象)到该集合的映射来刻画。例如，集合 $A=\{a,b,c\}$ 的元素有 3 个，分别为 a，b，c。相应地，单点集(终对象)到 A 的映射分别记为 $\bar{a}$，$\bar{b}$，$\bar{c}$，与集合里的元素一一对应。这样，就可以从映射的视角看待元素，记作 $\bar{a}\in_{\{*\}}A$，这里的单点集{ * }是元素 $\bar{a}$ 的上下文。由此推广得出集合元素的“新”概念——广义的集合元素，其个数取决于上下文到该集合映射的个数。例如，上下文由单点集改为含有两个元素的集合{ *，• }，此时由映射决定的 A 的元素个数为 9，不再是 3。这样，通过改变上下文，集合元素个数也会随之改变。若该上下文是空集，空集到 A 只有唯一的空映射，此时集合 A 中的元素就是 1 个。

根据对集合元素的“新”理解，我们可以给出悖论的一种新解释。

考虑增加了 $Cont$ 类型的布尔代数($bool$，$Cont$，T，F，$\neg$，$\vee$，$=$)，带有新函数 C:$Cont\to bool$。其满足公理：

$\neg(C(x))=C(x)$:$bool[x$:$Cont]$

经典情况下，$Cont$ 默认为单点集，记作 $\neg(C(x))=C(x)$:$bool[\ ]$，函数 C 表示为单点集到 $bool$ 类的映射，此时映射有两种，$bool$ 类中就有两个不同的元素 T 和 F，可区分判断 $C(x)$ 与 $\neg C(x)$，若满足上述公理，就出现 $T=F$:$bool[\ ]$的情形，从而导致悖论。除非 $Cont$ 类型为空，C 为唯一的空映射，$bool$ 类中有唯一元素，此时 $C(x)=\neg C(x)$:$bool[\varnothing]$成立，但只是平凡的情况。

以下借助时序集合，在不同阶段下通过改变上下文来构造出特殊的非空类型 $Cont$，使得每个时刻下函数 C 都具有唯一映射，即可满足上述公理，解释了该悖论。

考虑带有函数 E_X:$X\to\omega$ 的时序集合 X，ω 代表序数集合，函数 E_X 对 X 中的每个元素指派 ω 中的一个时刻。X 上有函数 $(-)^+$:$X\to X$ 代表集合 X 随时间的演化过程，满足 $E_X(x^+)=E_X(x)+1$。对于时序集合间的映射 f:$X\to Y$，满足时间不变性($E_Y(f(x))=E_X(x)$)以及演化不变性($f(x^+))=f(x)^+$。将 $bool$ 类型解释成在 0 时刻有两个不同的元素{T，F}，自 1 时刻以后演化成单点集的时序集。而 $Cont$ 类可看作一个在 0 时刻是空集，自 1 时刻以后是单点集的时序集。

对于函数 C:$Cont\to bool$，在 0 时刻，取 $Cont$ 为空类型，$bool$ 为{T，F}，则函数 C 有唯一的空映射，那么 $C(x)=\neg C(x)$:$bool[x$:$Cont]$成立。1 时刻开始，$Cont$ 类型演化为单点集，$bool$ 类型也演化为单点集，函数 C:$Cont\to bool$ 同样只有唯一映射，那么 $C(x)=\neg C(x)$:$bool[x$:$Cont]$成立。此时，$Cont$ 类型不是平凡的(空类型)。

根据集合元素的“新”解释，借助时序集合，构造出“新”类型 $Cont$，满足该公理 $C(x)=\neg C(x)$:$bool[x$:$Cont]$，悖论得以解释。

结束语　两次对角化过程必然导致悖论，对角线方法是探究自指代悖论生成机制的有力工具。这类悖论的出现是因为在系统内部涉及自指代问题都会出现不一致，简单禁止自指代实际上是一种消极的避免悖论的方法。

基于新视角，积极地看待悖论，从量子、范畴的角度来重新对悖论进行解释，给出了容纳悖论的模型。从量子力学视角，说谎者悖论作为一个认知实体，处于亦真亦假的叠加态，测量会改变语句原来的状态。在范畴理论领域，利用上下文获得推广的集合元素的新概念，构造出随时间演化的集合值函子(时序集合)，建立了容纳悖论的模型，进一步深化了我们对悖论本质的认识。

随着对悖论的深入认识，非自指代悖论——Yablo 悖论及 Curry 悖论将是以后探究的内容。

参 考 文 献

[1] 千红. 罗素悖论与数学危机[J]. 中国科技纵横，2002(10)：138-140.

[2] MARCHAL B. Theoretical computer science and the natural sciences[J]. Physics of Life Reviews，2005，2(4)：251-289.

[3] 张建军. 逻辑悖论研究引论[M]. 北京：人民出版社，2014：.

[4] KRIPKE S A. Outline of a Theory of Truth[J]. Journal of Philosophy，1975，72(19)：690-716.

[5] YANOFSKY N S. Computability and Complexity of Categorical Structures[OL]. http://www.researchgate.net/publication/280243417_Computability_and_Complexity_of_Categorical_Structures.

[6] YANOFSKY N S. A Universal Approach to Self - Referential Paradoxes，Incompleteness and Fixed Points[J]. Bulletin of Symbolic Logic，2003，9(3)：362-386.

[7] YUKALOV V I，SORNETTE D. Mathematical basis of quantum decision theory[J]. Swiss Finance Institute Research Paper，2008(08-25)：1-37.

[8] AERTS D，GABORA L，SOZZO S，et al. Quantum Structure in Cognition：Fundamentals and Applications[J]. Computer Science，2011，53(5)：314-348.

[9] YABLO S. Paradox without Self-Reference[J]. Analysis，1993，53(4)：251-252.

[10] PRIEST G. The Structure of the Paradoxes of Self-Reference[J]. Mind，1994，103(409)：25-34.

[11] SMITH N. The principle of uniform solution (of the paradoxes of self-reference)[J]. Mind，2000，109(433)：117-122.

[12] PARENT T. Paradox with just self-reference[J/OL]. http://www.unc.edu/~tparent/d.pdf.

[13] 杨义川，王拥军.《数理逻辑与集合论》中的对角化原则[J]. 大学数学，2017，33(1)：109-113.

[14] ABRAMSKY S，TZEVELEKOS N. Introduction to Categories and Categorical Logic[M]//Introduction to higher order categorical logic. Cambridge University Press，2011：3-94.