“知道者悖论”的博弈论分析

任晓明 谷飙

（中山大学逻辑与认知研究所，广州 510275;南开大学哲学系,天津，300071）

摘要:“知道者悖论”既涉及认知推理，也涉及主体间的策略性互动。本文以“突击考试”为例，指出这个悖论涉及合理行动问题，其中的推理是策略性推理，通过分析主体的目标、偏好、行动，在完全信息博弈模型中重构了相关推理，提出了一个博弈论解悖方案。

关键词: 知道者悖论 策略性互动 博弈论

[中图分类号] B81 [文献标识码] A

自20世纪50年代以来，“知道者悖论”一直受到哲学家和逻辑学家的关注。奎因(Quine）等人认为“知道者悖论”是逻辑谬误，试图通过分析悖论中的认知推理，找到推理的漏洞从而解决难题消除悖论。肖（Show)、蒙塔古（Montague）以及卡普兰（Kaplan）则认为“知道者悖论”中认知语句是自我指涉的，在形式语言中重构自我指涉就可使之成为严格意义上的逻辑悖论。 谬误说和悖论说只考虑了相关的认知语句和认知推理的语形和语义特征，忽视了其中的语用学要素：预设、语旨和语力,忽视了认知、理性和行动三者之间的关联。我们发现，悖论中认知概念和推理不只具有单纯的认知意义，而是与相关主体的目的、行动密切相关的，它们不是主体-客体形式的对象性认知，而是主体之间的互认知，是具有博弈思维特征的策略性互动。根据这一线索，我们针对“知道者悖论”的一个版本，提出一种不同于以往解悖方案的博弈论解悖方案。

一. “突击考试”难题

“知道者悖论”有许多版本,其中的主角可以是法官和死囚,国王和求婚者,或者老师和学生。下面的表述采用最后一个版本---“突击考试”难题。

星期天老师对学生宣称:（甲）下周周一到周五的五天内有且仅有一天下午举行考试,（乙）我保证在考试的当天上午你们不知道下午是否举行考试。

有一个聪明的学生做如下推理:

(1) 考试不可能安排在周五。假设考试安排在周五,则到周五的上午,我已确知在周一到周四的四天里没有考试,而考试一定安排在周一到周五的五天内,所以我可以肯定周五下午一定安排考试,即我已经在周五的上午知道下午会有考试,这与老师的保证“在考试的当天上午你们不知道下午是否举行考试”相矛盾。所以考试不可能安排在周五下午;

[**收稿日期**]2006-10-16.

[**基金项目**] 本文为教育部人文社会科学重点研究基地 2002-2003 年度重大项目《逻辑学在人文科学中的应用》（02JAZJD720018）的序列成果之一.

[**作者简介**] 任晓明（1953—— ），男，四川泸州人，南开大学哲学系教授，博士生导师，中山大学逻辑与认知研究所兼职研究员，主要从事逻辑学、西方科学哲学的研究。 谷彪（1967— ）男，四川营山人，南开大学博士研究生，主要从事逻辑学和科学哲学研究。

(2) 考试也不可能安排在周四。假设考试安排在周四,则在周四上午我已确知周一、周二和周三没有安排考试,所以考试只能安排在周四和周五。但是我在(1) 中已证明周五不能安排考试,所以考试一定安排在周四。这样我已在周四上午知道周四一定安排考试,这与老师的保证相矛盾,所以周四不能安排考试;

(3) 同理可证,周三、周二和周一也不能安排考试;

(4) 综上所述,本周根本不可能安排考试。

学生对自己的推理非常得意。可是在星期四的下午他大吃一惊,老师确实在周四下午安排了考试,而且他当天上午确实不知道下午将安排考试，因为根据他的推理，这一周的每一天都不可能安排考试。这样，“突击考试”不折不扣地实施了。

在上述“归谬法”论证中，(1) 无疑是最重要的。如果这个推理是正确的,那么(2) 、(3) 、(4) 也是正确的。即只要承认了周五不可以安排考试,接下来就必须依次承认周四、周三、周二和周一都不能安排考试。反之，如果(1)存在逻辑谬误,那么(2) 、(3) 、(4) 也不可能成立。因此，对悖论的分析自然要从(1)开始：在(1)中，学生的推理（如果我已确知在周一到周四的四天里没有考试,而考试一定安排在周一到周五的五天内,所以我可以肯定周五下午一定安排考试）依赖于对命题（甲） “本周一到周五的五天内有且仅有一天下午举行考试” 的断定。可是，后面的归谬论证却得出了“本周一到周五的五天内不会举行考试”这一结论 。这里的关键问题是：在周五的上午学生有没有可靠的依据断定（甲）？如果学生在在周五的上午不能断定（甲），这个推理还成立吗？

二．两种回答----谬误说和悖论说

奎因（Quine）等人认为学生推理中有两个预设：（a）学生知道“本周一到周五的五天内将有且仅有一天下午举行考试”是真的。(b）学生知道“在考试的当天上午你们不知道下午是否举行考试” 是真的。

奎因指出，预设（a）是不合理的，因为学生在做推理时没有可靠的依据断定“本周一到周五的五天内将有且仅有一天下午举行考试”是真的，而且，在他得出了“本周一到周五的五天内不会举行考试”这一结论后，如果在周五的上午已知周一到周四的四天里没有考试,他就应当怀疑而不是相信 “本周一到周五的五天内将有一天下午举行考试”。可以想象这样的情形：悖论中的承诺不是老师亲自对学生宣布的，而是由其他人传达的。随着时间的推移，学生发现在周一到周四的四天里没有考试。到了周五的上午，学生还会相信 “本周一到周五的五天内将有一天下午举行考试”吗？如果学生在周五的上午怀疑“本周一到周五的五天内将有一天下午举行考试”是真的，则老师在周五的下午安排考试是合理的，因为这时通知中的（甲）和（乙）都可以满足。因此，周五下午并不能从“突击考试”的可实施日期中排除，学生推理中“归谬法”的基础不复存在。奎因因此认为，不合理的预设导致学生做出似是而非的推理，取消了预设就解决了难题。他的结论是：“知道者悖论”不是悖论，只是一个逻辑谬误。

与奎因的谬误说相反，肖、蒙塔古以及卡普兰认为，“知道者悖论”中认知语句是自我指涉的，如果对原先的表述做修正，对认知概念及认知推理做模态化处理，就可以在形式语言中重构“知道者悖论”中的推理，形成一个严格意义上的逻辑悖论。

肖指出,在老师的承诺中“学生不知道在周一到周五举行考试”的真实含义应为“学生不知道根据我的承诺周一到周五将举行考试”。这个语句本身出现在老师的承诺中,而它的内部涉及了老师的“承诺”,因此实际存在着自我指涉。蒙塔古和卡普兰指出，在老师的承诺中加上一个选言支就可以构造一个严格的逻辑悖论。即，老师的承诺应修正为：

除非学生事先知道本承诺为假，否则以下要求之一将被满足：

或者(1) 本周仅周一安排一次考试,而学生在周日不知道基于本保证周一安排考试;或者(2) 本周仅周二安排一次考试,而学生在周一不知道基于本保证周二安排考试;或者(3) ..;或者(5) 本周仅周五安排一次考试,而学生在周四不知道基于本保证周五安排考试。

蒙塔古和卡普兰以这个表述为基础，在形式语言中对认知概念及认知推理做模态化处理，重构了老师的承诺和学生的推理，建立了两个命题之间的矛盾等价式，使“知道者悖论”成为一个严格意义上的逻辑悖论。

谬误说和悖论说有一个共同特征：从“知道”一词的认知意义出发讨论推理的语形和语义特征。例如，对学生推理的结论“本周根本不可能安排考试”，谬误说将它解释为学生不知道“突击考试在本周实施”是真的；悖论说将它解释为学生知道“突击考试在本周实施”是假的。从语用学的角度看，这个结论指的是“老师不可能在本周实施突击考试”，它涉及到行动的主体、行动能力、主体的目标、偏好，两个主体（老师和学生）间的互动。单纯的语形和语义分析，不可能阐释这个语句的预设、语旨和语力，更不可能给出“突击考试难题”的解答。在这个意义上，谬误说未能消解悖论，悖论说也未能使“知道者悖论”成为完整的语用悖论——它仍然是语义悖论的附属品。

三．目标、行动和认知——语用分析

在我们看来，“突击考试”中老师的考试通知并不是描述未来可能发生的事件，而是做出承诺并表明偏好，是一种以言行事的言语行为。老师的考试通知中，“本周一到周五的五天内将有且仅有一天下午举行考试”说明了老师在未来一周中的行动——要采用考试的形式促进学生的学习，“我保证在考试的当天上午你们不知道下午是否进行考试”表明了老师的偏好——希望学生在考试之前的每个上午都做考试准备，不希望学生在确知考试日期后在考试当天上午临时准备。对于老师而言，这个通知的合理性依赖于考试是否能够实施以及考试的实施能否取得他所偏好的结果。关键是：他的承诺和偏好能否相互协调？如果要兑现承诺，是否会使他偏好的结果无法实现？如果要实现偏好的结果，是否会使他的承诺无法兑现？显然，如果承诺和偏好不能相互协调，无论是承诺无法兑现还是偏好的结果无法实现，都会给老师带来损失。但是，最后的结果并不是由老师单方面的行动决定的，它也取决于学生方面的行动选择。在双方的互动中，老师可以通过对学生考试准备的观察和对学生推理的分析选择自己的行动，同时，老师知道学生也会去观察老师的行动并分析老师的目标和偏好从而选择自己的行动。正是从这种策略性互动的角度出发，老师在考试通知中明确了自己的行动选项和偏好，并且相信，双方根据这个通知选择行动，自己既能够兑现承诺又能实现偏好。

从学生方面看，他们偏好有准备的考试，不希望无准备的考试。那些打算在考试准备中投机取巧——仅在考试的当天做准备并能在考试中取得好成绩的学生是 “聪明”的学生，是追求边际效用最大化的理性行动者，对他们来说考试的当天上午做准备的边际效用大于考试之前的每天上午做准备，但考试的当天上午没有准备的效用是最低的。“聪明”的学生做出的那番推理实际上是一个选择最优行动的策略推理，他之所以认为老师承诺的突击考试不可能实施，恰恰是因为这种结果是他所偏好的：首先，他不用准备考试，无须付出成本；其次，与为无法实施的考试做准备的学生相比，不做准备自然是“聪明”的学生。换句话说，学生也是从策略性互动的角度看待考试通知，并且把老师看作博弈的对手。这就可以解释学生为什么对自己的推理很得意：他知道老师不实施已承诺的考试会带来声誉损失，因此在推理中他先将老师所承诺的“周一至周五有且只有一次考试”当成前提得出“如果在周一到周

四的四天里没有考试,考试一定安排在周周五”。然后以老师是理性的不会采用“周五考试”这一明显背离偏好的策略为依据，得出“周五不可能实施突击考试”的结论。总之，他认为，根据这个通知，老师无论怎么选择都不能使承诺和偏好相互协调，自己不做准备总是最优的策略。

然而，如果从策略性互动的角度看，老师实际上有六个可行行动：周一考试、周二考试、周三考试、周四考试、周五考试、不考试。学生之所以将“不考试”排除在外，根本原因在于他认为老师不实施已承诺的考试是非理性的。学生的偏好顺序是（在考试当天上午未准备但在之前的每个上午都准备）＜（在考试之前的每个上午都准备）＜（考试当天上午准备但在此之前的每个上午都不准备）。学生之所以相信“在考试的当天上午学生不知道下午是否进行考试”是真的，也是由于涉及合理选择相关行动：学生在知道下午进行考试的情况下会做准备从而实现自己偏好的结果，这与老师的偏好相反，因此老师肯定不会在学生能够知道下午进行考试的日期安排考试。因此，学生推理时不怀疑老师的通知的根本理由在于他相信老师不会在策略性互动中采用非理性的或明显背离偏好的行动。在得出“突击考试不可能实施”的结论后，他所否定的也不是老师通知的真实性，而是“突击考试”的可实施性。

四.“突击考试”博弈

根据以上分析，老师和学生的可行行动和他们对可能结果的偏好形成了博弈。其中，老师的可行行动是选择周一至周五的某个下午考试或不实施考试,学生的可行行动是选择在考试未到来的每个上午是否做考试准备。老师的偏好顺序是（学生在考试当天上午未准备但在之前的每个上午都准备）＞（学生在考试之前的每个上午都准备）＞（学生在考试当天上午准备但在此之前的每个上午都不准备）；学生的偏好与老师的偏好顺序相反。

我们注意到，不仅学生知道前几天是否已经进行了考试，而且老师也完全清楚学生在考试之前的每个上午是否做了准备。因此，这是一个完全信息动态博弈。为了便于分析，我们将老师和学生对可能结果的偏好表示为效用，其方法如下：

（1）老师可以承诺在一定时期进行一次考试，如果没有把承诺兑现，他的支付为承诺涉及的时期长度的负值；

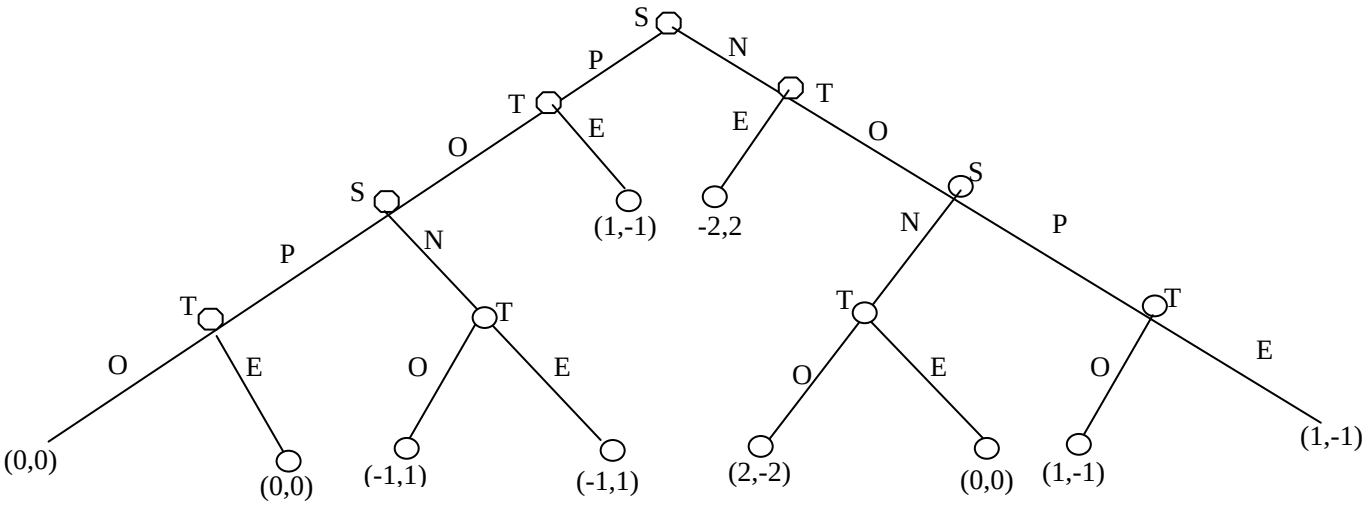
（2）学生可在承诺涉及的每一天的上午进行准备，每一次准备的成本为一个单位的效用值；

（3）如果学生在考试的当天上午准备，他的收益为考试的承诺期长度乘以一个单位的效用值；如果学生在考试的当天上午没有准备，他的收入为考试承诺期长度乘以一个单位的效用负值。

例如，老师承诺在未来 100 天里进行了一次考试，若考试在第 1 天进行而学生在当天上午做了准备，则博弈到达到一个终点，学生的效用值为 100-1=99；老师的效用值为-99；如果考试在第 100 天进行而学生在此之前每一天上午都准备了考试，则学生的效用值为 100-100=0，老师的效用值为 0；如果考试在 100 内没有实施而学生在此前的 n 个上午进行了准备，则学生的效用值为 100-n，老师的效用值为 n-100。

根据上述规则，博弈的长度是考试承诺涉及的时期长度的两倍，若老师只承诺“从明天开始的某天下午要进行一次考试”，则博弈是无限扩展博弈。若老师承诺“在周一下午进行一次考试”，则博弈的进程为两阶段，学生选择是否在周一上午准备考试，然后，老师选择是否选择在周一下午考试。这个悖论的实质内容同考试涉及的时期长度无关，因此，可以采用任意有限非负整数作为博弈的长度刻画老师和学生的认知推理和行动选择的合理性特征。下面我们假定老师承诺“在周一或周二下午有且仅有一次考试”，相应的博弈模型的树

形表示如下：



这个博弈有两个子博弈完美均衡（P,0,P,E），（P,0,P, 0)。(P,0,P,E）的含义是：学生在周一上午准备，老师未在周一下午考试；学生在周二上午准备，老师在周二下午考试。（P,0,P, 0）的含义是：学生在周一上午准备，老师未在周一下午考试；学生在周二上午准备，老师未在周二下午考试。也就是说，在这个博弈中，学生的理性行动是在考试前的每一天上午做准备，老师的理性行动是根据学生的准备情况选择考试时间。如果学生在某一天上午未准备，老师应当在这一天的下午考试；如果学生在每一天上午准备，老师可以在最后一天下午考试;如果老师不兑现考试承诺的效用损失可以由学生在每一天上午准备考试给他带来的效用补偿，老师也可以在最后一天下午不考试。

从博弈树可以看出均衡行动的算法：从最后一个决策点开始，找出决策者的最优行动，把它的结果当作该决策点的结果；然后，考察上一个决策点的最优行动，一直到初始点。不难发现，悖论中学生的推理与之相像。但是，学生只是分析了决策点上老师的行动，忘记了自己的行动对老师行动可能产生的影响。他在分析 “老师是否能在周五下午实施突击考试”时，没有注意到自己在周五上午有两个可行行动——准备考试或者不准备考试。如果不准备考试，老师就能在周五下午实施突击考试。学生推理的谬误正在于忽视了策略性互动，如果注意到自己的行动对老师行动可能产生的影响，正确推理的结论是，“如果我的策略是每个上午准备考试，则老师无法在本周实施突击考试”。显然，这个推理结论不会导致悖论。

综上所述，对知道者悖论，我们提出了一种博弈论解悖方案：根据语境中的预设，澄清相关主体的认知目标、偏好和可行的行动，构造他们之间策略性互动的博弈模型，运用均衡分析指出主体的行动和信念的可理性化特征，揭示了产生悖论的认知推理的谬误。

参考文献

[1] R.M.Sainsbury, 1995: *Paradoxes*, Cambridge University Press.

[2] 张建军，2002:《逻辑悖论研究引论》，南京：南京大学出版社。

.

GAME-THEORETIC ANALYSIS OF KNOWER PARADOX

Ren Xiao-ming Gu Biao

Abstract: knower paradox involves in cognitive reasoning and strategic interaction between distinct subjects. In this paper, unexpected exam is taken for example, it is pointed that knower paradox involves a rational action; its reasoning is strategic reasoning. It is reconstructed in a game model of perfect information by analyzing goad, prefer and action. A game-theoretic solution is proposed.

key words: knower paradox strategic interaction game theory