

doi:10.13582/j.cnki.1672-7835.2023.06.006

# 因果决策理论视野下的纽科姆难题

李莉, 李子毅

(湖北大学 哲学学院, 湖北 武汉 430062)

**摘要:**传统决策理论回避因果关系, 纽科姆难题的提出构成了对它的严重质疑。同时, 关于纽科姆难题的讨论推动了因果决策理论的进步。如吉伯德和哈珀基于因果关系的反事实理论为主观因果决策理论辩护; 斯科姆斯不使用反事实理论, 而在概率论基础上建构因果关系理论, 并为主观因果决策理论辩护; 索贝尔的决策理论是基于因果关系的概率性反事实理论; 刘易斯则基于模态逻辑提出了因果反事实的观点。这些理论的共同特点是使用因果关系来解决不确定性情况下的决策问题。然而基于决策逻辑的这种共同基因, 导致它们仍然存在各自的局限性。对这些问题的讨论, 有助于推进决策问题研究中关于概率、效用以及分区问题的思考。

**关键词:** 纽科姆难题; 因果决策; 概率; 反事实

**中图分类号:** B81      **文献标志码:** A      **文章编号:** 1672-7835(2023)06-0038-07

证据决策和因果决策是理性决策的两个原则与方法, 其理论依据都预设了理性人假定和期望效用最大化原则。根据这两个原则, 主体在一般决策问题上会得到一致答案。然而, 纽科姆难题的出现, 却让回避因果关系的传统证据决策理论受到了强烈质疑。支持因果决策原则的学者阐明了证据决策论由此陷入的困境, 同时给出了因果视域下的解决方案。这些方案之间既有联系也有区别, 例如, 尽管所有的方案都是基于因果关系, 但吉伯德和哈珀(Gibbard, A. and Harper, W. L.)<sup>①</sup>、索贝尔(Sobel, J. H.)<sup>②</sup>、刘易斯(Lewis, D.)<sup>③</sup>是基于因果关系的反事实理论, 而斯科姆斯<sup>④</sup>(Skyrms, B.)则不使用反事实理论, 致力于在概率论基础上建构因果关系理论, 并为主观因果决策理论辩护。吉伯德和哈珀为主观因果决策理论辩护, 而索贝尔的理论则是客观主义的。虽然存在这样的分歧, 但因果决策理论所呈现的相对于证据决策理论的优势, 使其迎来了因果理论在

决策研究中的复兴。而纽科姆难题作为对证据决策理论的关键反例, 也成为因果决策理论的“试剑石”, 进一步引发了决策问题研究中对相关概率、效用和分区的思考。

## 一 纽科姆难题的提出

纽科姆难题的提出者是利物浦实验室的物理学家纽科姆(Newcomb, W.)。诺齐克(Nozick, R.)以“纽科姆难题”为名在1969年发表了一篇著名的文章《纽科姆难题和选择的两个原则》。在该文中, 他把占主导地位的证据决策理论称为“标准理论”, 对纽科姆难题的作用作了如下阐释: “经济学家和统计学家已发展出一种有关合理决策的精心制作的理论, 并将其广泛运用到理论与政策研究之中。这是一种具有数学严格性的, 既强有力而又容易掌握的理论。虽然它作为实际行为的描述之充分性已受到广泛质疑, 它仍然是有关合理决策所应满足条件之研究中居于支

收稿日期: 2023-05-12

基金项目: 国家社会科学基金重点项目(23AZX018)

作者简介: 李莉(1974—), 女, 湖北鄂州人, 博士, 副教授, 主要从事逻辑哲学与逻辑应用研究。

<sup>①</sup>Gibbard A, Harper W L. “Counterfactuals and Two Kinds of Expected Utility”. Hooker C A, Leach J J, McClennen E F. *Foundations and Applications of Decision Theory*. Dordrecht Springer, 1978, pp.125-162.

<sup>②</sup>Sobel J H. “Circumstances and Dominance in a Causal Decision Theory”, *Synthese*, 1985, 63(2): 167-202.

<sup>③</sup>Lewis D. “Causal Decision Theory”, *Australasian Journal of Philosophy*, 1981, 59(1): 5-30.

<sup>④</sup>Skyrms B. “Causal Decision Theory”, *The Journal of Philosophy*, 1982, 79(11): 695-711.

配地位的标准理论。我认为,这种标准决策理论需扩充到与行为的符号意义及其他有关因素的明晰思考相结合。而关于当前标准理论之不充分性认识的一个有益入口,是由纽科姆难题提供的。”<sup>①</sup>文章发表后,引发了不同学科领域的许多学者讨论,在学界掀起了“纽科姆”热。纽科姆难题之所以能够持续引起人们的浓厚兴趣,主要在于它和其他的思想实验不同,它促使我们对理性选择的思考方式发生了深刻的转变。

纽科姆难题的最初版本来自于诺齐克的表述。他给出这样一个思想实验:(i) 拿走一个现在正面对着你的不透明盒子;或(ii) 同样的不透明盒子和旁边装有 1 000 美元的透明盒子。一个在预测人类在这种情况下行动方面有着良好的预测记录的超级生物,据此对你的选择进行了预测。如果它预测你只会选择不透明的盒子(单选),它就在不透明的盒子里放 100 万美元。如果它预测你会两个都选(二选),它就在不透明的盒子里什么也不放。对这个问题的解决主要可依据的原则有证据原则和因果原则:

证据原则是指一个理性的行为者会做最能证明他会实现其目标的事情;

因果原则是指一个理性的行为者会做他认为会使他实现其目标的事情。

根据证据原则,行为者应该选择单盒。因为拿走单盒,行为者就会获得百万美元,因此单盒是效用最大化的最好证据。而依据因果原则,行为者应该选择两盒。因为根据因果关系,不透明盒子中的内容已经确定了,行为者多拿走一个盒子可以保证额外的 1 000 美元。但问题是在这一决策情形中,证据原则和因果原则不可能都是对的。

## 二 为何需要因果关系

诺齐克提出纽科姆难题后,他本人认为这体现的是条件效用最大化原则和占优原则之间的冲突,并且当时并没有人把纽科姆难题跟因果决策理论关联起来,但后来在解决该难题的不断努力下,使得因果决策理论的重要性得以凸显。纽科姆难题有两个关键点,即预测者的预测和决策者的决策之间的概率依赖性,以及两者之间的因果独立性。对证据型决策理论来说,概率依赖性比

因果独立性更重要,而因果决策理论则赋予因果独立性比概率依赖性更多的权重。

所有传统的证据决策理论都是基于命题的偏好理论,如杰弗里(Jeffrey, R.C.)的概率主义及其决策运动学。即使是伊尔斯(Eells, E.)的共同原因理论,它也是对杰弗里决策逻辑的修订或改进。杰弗里本人不喜欢因果关系,他说:“证据决策理论的主要优点是,它不使用赌博的概念或任何其他因果概念。”<sup>②</sup>他用下面的例子说明了这一点。假设你的偏好排序包含命题 A“下周会有一场核战争”、B“下周会有好天气”和 C“这枚硬币正面向上”。那么考虑如下赌博情形:如果 C 发生,则 B;如果 C 不发生,则 A。即:如果这枚硬币正面向下,下周会有一场核战争;如果这枚硬币正面向上,下周会有好天气。根据杰弗里的说法,这种决策情形使得决策者并不能在他对命题的偏好排序中进行排序。此时如果假设这个赌博是有效的,那么决策者就必须修改他对战争和天气原因的信念,但问题是无法确定是哪些信念。此外,杰弗里更倾向于选择简单的理论。如果解释力一样,为什么不选择更简单的理论呢?当然,支持因果决策理论的人认为他们的理论比杰弗里的决策逻辑更适合解决复杂的决策问题。例如纽科姆难题的两盒方案论证中,对时间的直觉就体现了决策者对预测者的预测与决策者的决策之间的因果独立性的信念。同时,就分区依赖性而言,杰弗里决策逻辑是分区不变的,可以任意划分,而因果决策理论则是分区受限的。例如,如果决策者可以任意划分世界的可能状态,在纽科姆难题中,他也可以用以下方式划分世界的可能状态,如表 1 所示。

表 1 纽科姆难题中世界的可能状态

	S <sub>1</sub> :我在 T <sub>3</sub> 拿走 S <sub>2</sub> :我在 T <sub>3</sub> 拿走	
	两只盒子	盒子 2
a <sub>1</sub> :我在 T <sub>3</sub> 拿走两只盒子	O <sub>11</sub> :/	O <sub>12</sub> :/
a <sub>2</sub> :我在 T <sub>3</sub> 拿走盒子 2	O <sub>21</sub> :/	O <sub>22</sub> :/

这是纽科姆难题的决策矩阵,但其中没有任何的结果。这种对世界的可能状态的划分是没有意义的。因为从可能的行动与世界的可能状态的结合中没有产生任何可能的结果。纽科姆难题中对世界可能状态的无意义划分还有另一个例子,

<sup>①</sup>张建军:《逻辑悖论研究引论》,南京大学出版社 2002 年版,第 222 页。

<sup>②</sup>Jeffrey R.C. *The Logic of Decision*(second edition), The University of Chicago Press, 1983, p. 157.

是另外一种分区,如:我用左手拿箱子( $S_1$ )与我不是用左手拿箱子( $S_2$ )。

因此,因果决策理论支持者吉伯德和哈珀得出结论:一个可能的行动的效用不应该是衡量它作为可能状态的证据,而是衡量它在因果上带来一个想要的可能状态的功效。就纽科姆难题而言,虽然决策者会偏好拿走  $B_2$ ,这意味着他会得到 1 000 000 美元,但拿走  $B_2$  并不会因果地带来 1 000 000 美元。同样,尽管决策者不会偏好拿走两只盒子,因为这意味着有 0 美元,但拿走这两只盒子并不会因果地带来  $B_2$  中有 0 美元。然而,决策者应该拿走两只盒子。因为他知道,拿走这两只盒子会因果地带来一个想要的可能结果,即拥有额外的 1 000 美元,而他也知道同时拿走两只盒子不会带来他不想要的可能结果,即  $B_2$  中的 0 美元。

### 三 因果决策理论对纽科姆难题的消解

因果决策理论采用理性选择的原则,主要关注行为的后果。该理论认为,对理性选择的解释必须使用因果关系,从而确定使选择合理的因素。假如给定一组构成决策问题的选项,决策理论推荐一个效用最大化的选项,也就是说,一个效用等于或超过其他所有选项的选项。它通过计算一个选项的预期效用来评估该选项的效用,然后使用选项的可能结果出现的概率和效用来定义预期效用。由于这些概率依赖于选项,因果决策理论认为这种依赖性因果关系,而不仅仅是证据关系。

当代对因果关系的探讨有两个非常流行的版本,一个是因果关系的概率理论,另一个是因果关系的反事实理论。正如顿新国指出:“概率论因果理论的核心主张是原因改变(特别是提高)结果出现的概率。该主张的一个直接推论是,原因的出现并不必然伴随着结果的出现,反之亦然。”<sup>①</sup>也就是说,因果关系的概率理论说一个可能行动是一个可能结果的原因,是指它提高了该结果出现的概率。然而,如果要把因果关系的概率理论作为因果决策理论的基础,就必须确保因果独立性不能被简化为概率独立,否则就会使得纽科姆难题变成一个虚构的问题,不再具有任何意义。因果关系的反事实理论的基本思想是,因

果关系主张的含义可以用“如果 A 没有发生,那么 C 就不会发生”这样形式的反事实条件句来表达。大卫·刘易斯利用反事实条件句语义学首先给出了反事实因果关系的明确定义。接下来的问题是,因果决策理论中的那些因果关系理论是否为解决纽科姆难题提供了充分的基础?

所有因果决策理论在解决纽科姆难题上的共同特点是:(1)它们都把因果关系作为一个基本术语;(2)它们在问题解决上都归属于“两盒”解决方案。不同因果决策理论的区别,主要在于它们对可能行动的效用提出了不同的定义。

吉伯德和哈珀提出了可能行动的 U 效用定义,并建议决策者将 U 效用最大化:

$$U(a_i) = \sum_{j=1}^m c(a_i \square \rightarrow o_j) u(o_j) \quad (1)$$

其中  $a_i \square \rightarrow o_j$  意味着:“如果我做了  $a_i$ , 那么就会发生  $o_j$ 。”

斯科姆斯提出以下可能行动的 K 效用定义,并建议决策者将 K 效用最大化:

$$K(a_i) = \sum_{k=1}^1 \sum_{j=1}^m c(K_k) c(C_j | K_k \cap a_i) u(K_k \cap C_j \cap a_i) \quad (2)$$

其中  $a_i$  指可能行动,  $C_j$  指在决策时处于决策者因果影响范围内的因素,而  $K_k$  指决策时在决策者的因果影响之外的因素,但这些因素与决策者的可能结果有因果关系。

索贝尔提出了一个不同于吉伯德和哈珀的可能行动的效用定义,亦称 U 效用。他认为决策者应当采取 U 效用最大化原则。但索贝尔这里的 U 效用把机会(chance)或概率纳入了考量范围。对可能行动  $a_i$ :

$$U(a_i) = \sum_w \{ch(w/a_i)\} u(w) \quad (3)$$

刘易斯则提出了如下可能行动的效用的定义:

$$U(a_i) = \sum_{j=1}^m \sum_q c(a_i \square \rightarrow [P(s_j) = q]) qu(a_i \cap s_j) \quad (4)$$

其中,  $s_j$  表示任一分区, P 表示分区的概率分布函数, q 表示任一从 0 到 1 的数字。刘易斯声称,决策者应该使这个效用最大化。

鉴于上述异同,我们将这几大因果决策理论

<sup>①</sup>顿新国:《因果理论的概率论进路及其问题》,《哲学研究》2012 第 7 期。

作如下具体评述。

### (一) 基于虚拟条件的 U 效用

根据吉伯德和哈伯的观点,决策者应该采取最大化 U 效用原则,并对 U 效用进行了界定。决策者此时将可信度(C)赋予他在决策中考虑的虚拟条件。因为,在通常情况下,决策者并不确定如果他采取某种可能的行动会发生什么情况。因此,有上列 U 效用的计算公式。

吉伯德和哈伯将使用虚拟条件概率的因果决策理论与使用条件概率的证据决策理论区分开来。因为虚拟条件概率隐含了因果关系,据此可以计算选项的预期效用,从而使理论具有因果性。在他们看来,一个虚拟条件句的前件可以是假的,也就是说决策者可以相信形式为  $a_i \square \rightarrow s_j$  的虚拟条件句既代表他可能会执行的行动,也代表他可能不会执行的行动。如果  $a_i$  的成立因果地带来了  $s_j$  的成立,那么虚拟条件句  $a_i \square \rightarrow s_j$  是真的,仅当  $s_j$  成立。此时,条件概率就不适合用来计算期权的期望效用,因为相关的概率之和不等于 1(或者甚至不存在)。吉伯德和哈伯发展了斯塔内克及托马森(Stalnaker, R. and Thompson, R.)的可能世界语义理论,并给出了两个公理来支持这个理论。具体到纽科姆问题中,假设  $S_1$  的可信度为  $\mu$ ,  $S_2$  的可信度是  $1 - \mu$ 。因为  $S_1$  和  $S_2$  与决策者的可能行动  $a_1$  和  $a_2$  在因果关系上是独立的。因此决策者的可信度为:

$$c(a_1 \square \rightarrow s_1) = \mu$$

$$c(a_2 \square \rightarrow s_1) = \mu$$

$$c(a_1 \square \rightarrow s_2) = 1 - \mu$$

$$c(a_2 \square \rightarrow s_2) = 1 - \mu$$

对可能的行动的 U 效用计算得到:

$$\begin{aligned} U(a_1) &= c(a_1 \square \rightarrow s_1)u(1\ 000) + \\ &\quad c(a_1 \square \rightarrow s_2)u(1\ 001\ 000) = \\ &\quad 1\ 000\mu + (1 - \mu)1\ 001\ 000 = \\ &\quad 1\ 001\ 000 - 1\ 000\ 000\mu \end{aligned}$$

$$\begin{aligned} U(a_2) &= c(a_2 \square \rightarrow s_1)u(0) + \\ &\quad c(a_2 \square \rightarrow s_2)u(1\ 000\ 000) = \\ &\quad \mu(0) + (1 - \mu)1\ 000\ 000 = \\ &\quad 1\ 000\ 000 - 1\ 000\ 000\mu \end{aligned}$$

$U(a_1) > U(a_2)$ , 因此 U 效用最大化原则建议采取“两盒”决策。

吉伯德和哈伯还考虑了纽科姆难题的变体,其中已知预测者是不会错的。因此,对吉伯德和哈伯的批评主要集中在推理的前提方面,即“预

测者无懈可击”。显然,“无懈可击的预测者预测正确”这一命题在逻辑上是必然真的,但一个无懈可击的预测者在实际上是不可能存在的。此外,有学者批评他们并没有提出可接受的虚拟条件逻辑,并且没有为虚拟条件给出非标准概率理论。斯科姆斯指出,在吉伯德和哈伯的理论中,用于计算可能行动的 U 效用的虚拟条件必须被赋予非回溯性解释。这个批评的确是合理的,因为吉伯德和哈伯要求将虚拟条件诉诸直觉。但是有的人可能对虚拟条件有较弱的直觉,也有人有强的回溯直觉,因而吉伯德和哈伯的因果决策理论在应用上是有局限的。

### (二) 基于概率的 K 效用

斯科姆斯提出了一个不使用虚拟条件概率的因果决策理论。该理论所计算的行为的期望效用是该行为的 K 效用。斯科姆斯的因果决策理论在用“两盒”方案解决纽科姆难题上有自己的特色。首先,他的理论在最终分析中是完全主观的,不要求决策者的客观性。其次,他使用无条件的信任来计算一个可能的行动的效用,从而假设决策者对其可能的行动有完全的控制。再次,他不使用虚拟条件,因此他的理论也可以适用于对虚拟条件直觉较弱和/或有较强回溯直觉的决策者。最后,斯科姆斯的因果决策理论并不要求给出最终的可能结果,因此对决策者的知识要求较低。

根据斯科姆斯的说法,亚里士多德是第一个声称因果、模态或反事实的区别是理性决策理论的核心的人。决策者应该考虑的是他们可以通过可能的行动产生因果关系的因素,而不是已经确定不变的因素,比如说几何学或自然律等。因此,斯科姆斯以前列公式计算一个可能行动的 K 效用。

斯科姆斯关于 K 效用的计算方式看上去是合理的,但有一个遗留问题。如何判断哪些事实是在因果影响之外? 斯科姆斯辩护说,决策者无论如何都会构建一个依赖性假设,这些假设的真值就是在因果影响之外的因素。斯科姆斯声称,他的因果决策理论与其说是对斯塔内克反事实理论的推进,不如说是对萨维奇(Savage, L.J.)的理性决策理论的推进。他认为决策者不能决定世界的可能状态,但能决定可能的行动。于是在定义一个可能行动的 K 效用时采取两步走的策略。先计算出某个世界状态下某个可能行动的客观效用,然后一旦这个世界状态得以实现,决策者就可

以最高的客观效用为标准来决定采取何种行动。如果没有实现,就意味着它不知道哪种行动有最高的客观效用,此时就应当用最高的主观效用来确定行动。此时,就可以用无条件的置信度来定义主观效用。

斯科姆斯指出,在纽科姆问题中此时世界的可能状态是:( $K_1$ ) 0美元在  $B_2$ , 1 000美元在  $B_1$ ; ( $K_2$ ) 1 000 000美元在  $B_2$ , 1 000美元在  $B_1$ 。决策者可能采取的行动是:( $a_1$ ) 拿两个盒子; ( $a_2$ ) 拿  $B_2$ 。决策者的可能收益是  $C_{11}$ : 1 000美元;  $C_{12}$ : 0美元;  $C_{21}$ : 1 001 000美元;  $C_{22}$ : 1 000 000美元。因为如果决策者拿走  $B_2$  (其中有1 000 000美元), 得到1 000 000美元的机会是1, 于是有:  $U_{k1}(a_1) = 1 000$ ;  $U_{k1}(a_2) = 0$ ,  $U_{k2}(a_1) = 1 001 000$ ,  $U_{k2}(a_2) = 1 000 000$ 。由于机会要么是0要么是1, 所以在每个可能的状态下, 拿两个盒子的主观效用都高于只拿  $B_2$  的主观效用。因此建议决策者同时拿走两个盒子。

斯科姆斯使用无条件的信任来计算可能行动的主观效用, 是因为他认为决策者的可能行动并不因果地影响世界的可能状态。例如, 先排除了世界可能状态的分区: 预测者在  $t_1$  时预测正确 ( $S_1$ ) 与预测者在  $t_1$  时没有预测正确 ( $S_2$ )。主要原因是这些分区不符合他的要求, 因为世界的可能状态与决策者的可能状态是因果独立的。与吉伯德和哈珀的因果决策理论相比, 斯科姆斯的因果决策理论不使用虚拟条件, 也不要求以最终的可能结果来表述。斯科姆斯对期望效用的计算省去了因果概率, 转而在世界状态中建立了因果关系。在纽科姆问题的决策情形中, 这个计算结果与采用虚拟条件概率的期望效用的计算结果是相同的。

然而, 在认识论上对决策者的要求而言, 杰弗里的概率论要优于斯科姆斯的因果决策理论。斯科姆斯的理论预设了一个理想化的决策者, 在认识论上他对决策者提出了要求, 而杰弗里的概率决策论在这个意义上没有任何限制。尽管如此, 鉴于斯科姆斯理论的种种优点, 到目前为止它仍被认为是最充分的理性决策理论。

### (三) 基于概率性反事实的因果关系方案

索贝尔<sup>①</sup>则采取了一条不同的路径。他试图通过改进杰弗里的决策逻辑来建立自己的因果决

策理论。索贝尔认为杰弗里的决策逻辑在解决纽科姆难题上给出了错误的建议, 主要原因是杰弗里的理论没有体现因果关系。但是, 杰弗里的理论对所有可能状态的分区有效的特点一直吸引着索贝尔, 因此他通过对杰弗里的决策逻辑进行改造, 建立了自己的因果决策理论。

该理论由杰弗里的决策逻辑通过四个定义改造而来, 适用于世界可能状态的所有自然分区。首先, 索贝尔对效用的定义更接近杰弗里对效用的定义, 因此可以兼容杰弗里的决策逻辑。其次, 索贝尔的效用定义并不预设另一个可取的定义。再次, 索贝尔的效用定义使他的因果决策理论成为标准模态逻辑的自然延伸。从杰弗里的证据决策理论推导出索贝尔的因果决策理论的过程分为三个步骤, 并可以通过四个定义来实现。索贝尔没有使用条件置信度, 而是引入  $p \square \rightarrow q$  来衡量因果关系。他认为在效用计算中出现的概率应该是衡量因果潜力, 而不是衡量证据潜力。  $p \square \rightarrow q$  表达了因果条件, 但不完全是因果条件。

在纽科姆难题的认识上, 索贝尔与吉伯德和哈珀的观点一致, 认为纽科姆难题并不是条件效用最大化原则与强占优原则之间的冲突, 而是两个最大化原则, 即证据最大化原则和因果最大化原则之间的冲突。

首先, 索贝尔声称, 在纽科姆问题中, 世界可能状态有以下两个自然分区:

$S_1$ : 预测者在  $t_1$  时已经预测到我将取走两个盒子, 他在  $t_2$  时将 0 美元放入  $B_2$  中;  $S_2$ : 预测者已经预测到在  $t_1$  时, 我将接受  $B_2$  的内容, 并且他在  $t_2$  时在  $B_2$  中放入了 1 000 000 美元。

$S_1$ : 预测者在  $t_1$  时预测正确;  $S_2$ : 预测者在  $t_1$  时没有预测正确。

然而, 从经验的角度来看, 索贝尔认为选择拿走两只盒子的决策者倾向于第一种分区, 即因果分区; 而选择拿走  $B_2$  的决策者则倾向于第二种分区, 即证据分区。

对第一个分区, 索贝尔根据证据决策理论计算得出:

$$V(a1) = [0] 1 001 000 + [1] 1 000, = [1 000]$$

$$V(a2) = [1] 1 000 000 + [0] 0 =$$

<sup>①</sup>Sobel J H. "Notes on Decision Theory: Old Wine in New Bottles", *Australasian Journal of Philosophy*, 1986, 64(4): 407-437.

[1 000 000]

在这种情况下,证据决策理论建议取  $B_2$ 。

对第二个分区,索贝尔根据因果决策理论的计算结果是:

$$U(a_1) = (1 - p)1\,000 + p(1\,001\,000) = 1\,000 + p(1\,000\,000)$$

$$U(a_2) = (1 - p)0 + p(1\,000\,000)$$

显然,索贝尔的因果决策理论同样属于“两盒”方案。他通过放弃证据最大化原则和维护因果最大化原则来解决原来的冲突。索贝尔的因果决策理论是以世界可能状态的自然分区为基础的,因此,世界可能状态的两个分区,一个是因果的,一个是证据的。但是,问题在于自然世界可能状态的分区并不能等同于世界可能状态的理性分区,所以这一点也是他的理论受到批评最多的地方。

#### (四) 基于反事实条件句的因果关系

刘易斯的因果决策理论基于两个论点:第一,决策制定者应通过依赖性假设计算  $U$  效用最大化。第二,依赖性假设是全概率模式的结合物。刘易斯的因果决策理论可以被看作是对吉伯德和哈珀的因果决策理论的推进,并建基在他自己的反事实条件句的可能世界语义学之上。他认为,决策者的信念和愿望由决策者的置信度和效用函数来表示,这些函数是在单一的可能世界上定义的。对任一世界  $w$  而言,  $C(w)$  表示决策者认为  $w$  为实际世界的置信度。这个置信度范围从 0 到 1,且和为 1。每个世界  $w$  都有一个效用  $u(w)$  衡量决策者对  $w$  为真实世界的满意程度。他指出一个分区是一组命题,其中有一个命题在任何可能世界都成立。据此,他提出了自己的因果决策理论。该理论使用依赖性假设的概率而不是虚拟条件的概率来计算预期效用。

刘易斯对吉伯德和哈珀的因果决策理论是这样描述的:在日常语言中,依赖性假设通常由反事实条件句来表达。例如,假设决策者希望猫咪布鲁斯能够发出呼噜声,决策者的选择是刷牙、抚摸和独自离开。而猫咪布鲁斯则可以大声地、轻声地或者根本不打呼噜,这些选择建立了一个丰富的分区。那么,大部分决策者对以下依赖性假设的置信可由三个反事实条件句来表示:

(1) 我给布鲁斯梳毛  $\square \rightarrow$  他大声叫唤;

(2) 我抚摸布鲁斯  $\square \rightarrow$  他轻声

叫唤;

(3) 我让布鲁斯一个人呆着  $\square \rightarrow$

他就不打呼噜。

这一依赖性假设指出,轻柔和响亮的呼噜声在因果上取决于决策者的选择。此外,它还规定了决策者的影响程度,即完全控制,并规定了决策者影响的方向,即他必须做什么才能得到什么。但决策者也对其他依赖性假设给予一定的置信,例如对以下情况:

(1) 我给布鲁斯梳毛  $\square \rightarrow$  他不打呼噜;

(2) 我抚摸布鲁斯  $\square \rightarrow$  他不打呼噜;

(3) 我让布鲁斯一个人呆着  $\square \rightarrow$  他不打呼噜。

这种依赖性假设指出,不打呼噜与决策者的选择没有因果关系。将两个依赖性假设相比较,就会发现其实是反事实的模式而不是一个反事实条件句所单独表达的因果依赖性或因因果独立性。

总之,刘易斯认为他的理性决策理论在两个不同的方面是因果理论。一方面,依赖性假设的内容是因果的;另一方面,依赖性假设本身在因果关系上与决策者的可能行动无关。根据刘易斯的说法,正确的划分应该是对依赖性假设的划分,强调的是其因果内容而不是其因果独立性。刘易斯的因果决策理论是对吉伯德和哈珀决策理论的推进。他指出吉伯德和哈珀的反事实条件是受限的,因为此时前因和后果的反事实性必须指定完全不同的事件,这些事件既要引起前因,也要引起后果。此外,回溯性的反事实条件被排除在外。

在刘易斯看来,吉伯德和哈珀是这样建构他们的因果理论的:首先,吉伯德和哈珀认为因果反事实的形式为形式的反事实,即  $a_i \square \rightarrow s_j$ 。对每个选项,他们将一个完整的模式定义为由形式为  $a_i \square \rightarrow s_j$  的反事实条件句组成的集合。其次,在一个完整的模式中,反事实条件句的组合形成了吉伯德和哈珀式的依赖假设。最后,假设每个世界都有一个完整的模式,这样就完成了吉伯德和哈珀的因果理论。对这个决策理论,刘易斯提出两个反对意见,任意性反对和概率反对。就概率反对意见来看,假设决策者相信实际的世界可能是不确定的,许多事情是由概率过程解决的,那么,决策者可能会对完全模式成立的世界或形式为  $a_i$  的反事实给予很少的信任。 $s_j$  构成这些完整

模式的反事实。因此,吉伯德和哈珀将依赖性假设视为完整模式的联结点的概念过于狭隘。刘易斯在此基础上,建立了自己的因果决策理论。他认为应该将置信分配给或然的具有单一情况的命题。例如,刘易斯给出了以下三个概率命题:

[ $P = p_1$ ]: 布鲁斯大声打呼噜的概率是 50%;他轻声打呼噜的概率是 40%;而他完全不打呼噜的概率是 10%。

[ $P = p_2$ ]: 布鲁斯大声打呼噜的概率是 30%;他轻声打呼噜的概率是 50%;而他完全不打呼噜的概率是 20%。

[ $P = p_3$ ]: 布鲁斯大声打呼噜的概率是 10%;他轻声打呼噜的概率是 10%;而他完全不打呼噜的概率是 80%。

这个例子是一个全概率模式中的因果反事实的结合:

我给布鲁斯梳毛 [ $P = p_1$ ] 成立;

我抚摸布鲁斯 [ $P = p_2$ ] 成立;

我让布鲁斯一个人呆着 [ $P = p_3$ ] 成立。

这个因果关系假设就可以表达大声呼噜和小声呼噜对决策的影响。此外,还可表达决策者的影响方向。所以吉伯德和哈珀的依赖关系假设其实是刘易斯依赖性假设的一个极端情况。总的说

来,相比吉伯德和哈珀的因果决策理论,刘易斯的理论具有更广泛的适用性。

### 结语

传统的证据决策理论反对在决策中使用因果关系,但是在纽科姆问题的决策情形中,因果决策理论似乎给出了更令人满意的答案。各种因果决策理论都使用了因果关系,但是却在不同层面上具有各自的特点。很难说哪一个更出色地解决了纽科姆难题,但每种理论不同的侧重点都告诉我们因果关系在决策当中不可或缺。伴随着不同的观点,因果决策理论已经渗透到哲学的各个领域,它依赖于形而上学对因果关系的解释,同时也依靠归纳逻辑来说明有关因果关系的推论。一个全面的因果决策理论,必须进一步回答决策问题中的概率和效用问题。至于纽科姆难题自身,它不仅是当代决策理论研究的一个关键案例,也具有跨学科研究的意义,在现实生活中的各个领域都存在着纽科姆难题的影子。揭示纽科姆难题及其解决方案中蕴含的哲学意蕴,特别是纽科姆难题背后的因果关系,将有助于彰显此类决策悖论与当代行动理论和行动哲学研究的深层关联,推动认知科学、认知逻辑和人工智能研究的发展。

## Newcomb's Problem from the Perspective of Causal Decision Theory

LI Li & LI Ziyi

(School of Philosophy, Hubei University, Wuhan 430062, China)

**Abstract:** Traditional decision theory avoids causation, but the Newcomb's problem constitutes a counter example to traditional decision theory, and the discussion of the Newcomb's problem has contributed to the advancement of causal decision theory. For example, Gibbard and Harper defended the subjective causal decision theory, which is a counterfactual theory based on causality. Skyrms's theory of causality based on probabilistic theories defended the subjective causal decision theory. Sobel's theory is a probabilistic counterfactual theory based on causality, and Lewis put forward a causal counterfactual view based on modal logic. The common feature of these theories is the use of causality to solve decision problems under uncertainty. However, this common genesis based on decision logic leads to the fact that they still have their own limitations. The discussion of these issues will further provoke thinking about probability and utility issues in decision problems and about partitioning.

**Key words:** Newcomb's problem; causal decision making; probability; counterfactual

(责任校对 唐尧)