逻辑学研究　2025 年第 2 期，1–23
文章编号：1674-3202(2025)-02-0001-23

条件化、干预和影像
——信念度改变的三径路

吴小安

摘　要：本文讨论了信念度改变的三种方式。前两种是在获得新的观测信息或者干预信息的情况下，一个完美的理性主体如何改变信念度的方法；后一种是关于在可能世界的框架内如何对反事实指派概率的影像理论，同样也是一种关于概率如何流变的理论。我将说明和比较在可能世界的框架如何来理解这三种信念度的改变。珀尔将以干预理论的运作机制为参考，修正刘易斯的影像理论，这个修正的影像理论相较于干预理论有其优势，比如对于析取前件反事实的处理，但我将证明因为这个修正要把刘易斯原先所采用的罗伯特 • 斯塔内克语义替换为他自己的语义，不再能得到原初的结果，即 $\mathbf{P}^{A}(C) = \mathbf{P}(A \boxdot\!\!\rightarrow C)$。

关键词：条件概率；影像；结构因果模型；加权系统

中图分类号：B81　　**文献标识码：**A

1　导论

反事实推理的应用非常广泛，尤其在统计学、社会科学和法学等领域得到了广泛讨论。对反事实条件句语义的哲学研究也因此自然而然地引起了关注。最为人知的当属大卫 • 刘易斯（D. Lewis）和罗伯特 • 斯托奈克（R. Stalnaker）提出的反事实可能世界语义理论。该理论的基本思想是：形如“If A were the case, then B would be the case”的反事实条件句在世界 w 中为真，当且仅当 B 在所有最接近 w 的 A-世界中为真。而世界之间的“接近性”由它们之间的“相似性”决定。显然，如果要用该语义来判定具体反事实的真值，就需要一个能够加权判断可能世界之间相似性的加权系统。

收稿日期：2024-12-03
作者信息：吴小安　西北工业大学马克思主义学院
wuxiaoan1984@126.com
基金项目：国家社科基金重大项目“西方语言哲学前沿问题研究”（23&ZD240）。

进入 21 世纪后，朱迪亚·珀尔（J. Pearl）提出基于结构方程的因果模型来进行因果推理。因果模型本质上也是一种通过反事实来分析因果关系的理论。（[14, 15]）在因果模型中，反事实通常涉及预测干预（intervention）的效应，回答诸如“If random variable X were set to x, what would the value of random variable Y be”这样的问题。因果模型所讨论的反事实通常是相对有限的一类，通常将其称为“结构反事实（structural counterfactuals）”。这也意味着，因果模型能够为（至少是某些）反事实提供语义（即真值条件）。尽管两者从不同的视角切入问题，但它们都需要处理一个基本的共同问题。若将因果模型中所有变量的一次取值理解为一个世界，便可以发现结构路径和可能世界路径之间在某些方面的相似性与差异性，通过它们各自的优缺点碰撞，深入理解它们的内在含义。基于此，本文将讨论在获得新信息后，一个理性主体如何通过三种方式调整其信念度。在具体阐述这些路径之前，本文将对可能世界语义和因果模型框架进行必要的比较和介绍，以便更好地理解后续的讨论。

2 结构因果模型和反事实的可能世界理论

粗看之下，因果模型语义和可能世界语义实在太不一样了。首先，领域不同。因果模型是要从大数据中发现因果关系，是社会科学工作者的方法论和工具，而反事实的可能世界逻辑则是逻辑学家和哲学家的主场，正如经典逻辑并不对原子句的实际真值很关切一样，发展反事实逻辑的初衷并不是想用它来讨论具体反事实的真值条件。对真值条件的讨论只要能够满足解决逻辑的问题就好，比如 MP 规则。尽管刘易斯给出了一个判定可能世界之间相似性的加权系统（[13]），但是“太含糊，也太不清楚了，以至于不能为大多数科学语境中的反事实判定提供指引。”（[21]，第 168 页）

其次，两者所处理的问题也不同。因果模型为应用而生，它的文献中处理的问题一般如下：“吸烟是否导致肺癌？”，“一个特定的疗法是否对预防某类疾病有效？”，“是新的税法还是层出不穷的广告推销导致了产品销售额的增长？”所以不管理论本身或者数学上面如何漂亮，如果对于现实问题的解决没有裨益，那么终究会被“扬弃”。而可能世界框架则是“为美而美”的存在，或者说是为逻辑的美而生，它从人们是如何判定反事实的基本直觉为出发，以可能世界框架来构建出恢弘的思想大厦，它的主业并不在于判定具体反事实的真值，作为它们理论阐述的出发点和引子的那些例子：“如果袋鼠没有尾巴，那么它就会跌倒”，“如果 Oswald 没有刺杀 Kennedy，那么其他人将会刺杀他”清晰直白，引用这些例子更多是为了来佐证理论设想本身的合理性。

再次，两者所努力的最终目标也不一样。对于逻辑学家而言，他们更关心的是语义系统以及与之对应的公理系统是否满足完备性和可靠性。且自然语言中的反事实的具有一些逻辑特征，比如，不能加强前件（No augmentation），不能不能质位互换（No contraposition），那么所给出的反事实的逻辑系统是否也具有这些特征呢？如果自然语言中有效的推理模型不被逻辑系统所含括，如果日常语言中无效的推理模型（比如前件增加）不被系统所排除，那么应当如何来修正公理系统以及语义约束以得出与直觉推理相符合的结论呢？这是大部分的逻辑工作孜孜努力之处。（[1, 4]）所以一个简单析取前件（Simplification of Disjunctive Antecedents，SDA）和**经典等价式的代入**问题就从 1975 年一直争论到现在。（[3, 4]）

因果模型的终极目标是帮助未来的人工智能进行因果推理（[17]），在珀尔看来，因果问题的解决是实现强人工智能的关键一步。在珀尔看来人的优越性在于他能够回答为什么的问题，人的大脑是处理因果关系最为先进的工具，存储着海量的因果知识，这些知识是我们对周边环境的一种心灵表征，通过拷问这个表征，通过“想象”这种心灵能力来删改这个表征的某些部分，来回答各式各样的问题。这才是人工智能要努力了解的“人之为人”的这样一种超进化的能力。

尽管上述那么多差异，但还是能够从中找到一些相同点。首先，它们关于如何判定反事实的出发点和基本直觉是共通的。可能世界框架中的偏离的“小奇迹”和因果模型中的“完美干预”都是对使反事实前件为真但又不无端偏离现实世界要求的不同贯彻。其次，尽管可能世界框架宣称他们主要关心的并不是具体反事实的真值条件的判定，而是在给出语义的情况下的逻辑问题。

> 分析 2（刘易斯所给出的反事实语义的最终分析）（加上关于世界之间的相似性的形式特征的一些简单观察）穷尽了对反事实条件句所能够做的最一般的论断。尽管它并不是完全没有可检测其正确性的情境——它解决了逻辑的一些问题——它对于预测在特定情境中特定反事实条件句的真值贡献是很小的。（[13]，第 465 页）

但它们的理论还是涉及到一些上面所述的反事实条件句的真值判定。当然因果模型所计算出来的反事实条件句不是一个确切的真值，而是概率。这也符合常识，当判定一个反事实为真的时候，更多是因为觉得这个反事实成立的概率很大，而且有的反事实是无法判定真值的，比如“如果我抛那枚硬币，那么它将会人头朝上”，显然在可能世界的框架中，这个条件句为假，但是更多会觉得这个条件句有 50% 的概率为真。对于这样的问题，可能世界的框架也不是没有作为，刘易斯提出了影像（imaging）概念（[12]），所做的就是如何在可能世界的框架下，给反事实条件句指派概率。

最后，刘易斯强调他所用来分析因果的反事实是非-回溯的（non-backtracking）反事实，为了消解反事实本身的含糊性需要这样一种“标准解决”，他在哲学上系统的证明了反事实依赖的不对称。（[13]）而因果模型中的结构方程本身就暗含着一种不对称，方程右边的变量直接影响方程左边的变量，如果对结构方程做反事实的解读（[9]），那么因果模型用来分析因果的那类反事实同样也是非-回溯的，尽管它并没有觉得需要对这样一种不对称进行严格的哲学论证。

接下来将从一个服药-康复的例子出发来，考察在条件化和干预的不同处理之下，概率分布的流动演变（3.1 节和 3.2 节），在 3.3 节中，将证明刘易斯的影像理论没有给出相似性度量的标准，所以不能处理这些现实的数据情形，但是参考干预之下概率的流变过程之后，以之为依据给出相似性判定标准和松动影像理论前提预设之下，珀尔的这个修正影像理论不仅可以实现干预的等价效果，而且还可以处理结构因果理论所不能处理的析取前件反事实，但是修正影像理论也带来了新问题，不再满足 $\mathbf{P}^{A}(C) = \mathbf{P}(A \boxdot\!\!\rightarrow C)$ 。

3 条件化，干预和影像

因果模型和**影像**对于反事实的概率指派都是以经典的概率理论为基础的，且它们本质上都是**条件概率**的一种变体，所以要理解前两者，就要对后者的运作方式有所了解。通过一个具体的例子来说明条件化是有益的。这是一个关于病人服药与康复的例子。数据（1a）和因果图（1b）如下，注意这里假定了性别因素对于病人服药和康复有影响。其中 $X = 1$ 表示病人服药，否则就没有服药；$Y = 1$ 表示病人康复，否则病人就没有康复；$Z = 1$ 表示病人性别为男性，否则为女性。

有了上述数据以及与之对应的因果图[1]就可以考察在观测到病人服药（$X = 1$）的情况之下或者在干预病人服药（$do(X = 1)$）的情况之下，对于病人会康复的信念度的转变。让我们先考察条件概率的概率流变。

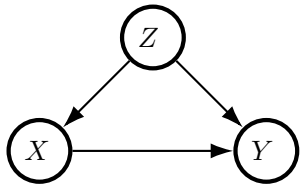
3.1 条件化

为了充分地展现在观测到 $X = 1$ 之下信念度的转变，考察条件概率 $\mathbf{P}(Y = y, Z = z \mid X = 1)$，它表示的是在病人服药的条件之下，$(Z = z \wedge Y = y)$ 的概率为何。根据条件概率和概率的公理，很自然地可以推导如下：

[1]我们预设的因果图表示的是数据背后的故事及对应数据的生成机制，对于如何通过数据来寻找，背后的因果图需要很强的假定。

	X	Y	Z	百分比
w_1	1	1	1	0.116
w_2	1	1	0	0.274
w_3	1	0	1	0.009
w_4	1	0	0	0.101
w_5	0	1	1	0.334
w_6	0	1	0	0.079
w_7	0	0	1	0.051
w_8	0	0	0	0.036

(a) $\mathbf{P}(X, Y, Z)$ 的联合分布



(b) 药物例子的因果图

图 1: 药物–康复例子的数据以及所对应的因果图

$$
\begin{aligned}
\mathbf{P}(Y=y, Z=z \mid X=1) &= \frac{\mathbf{P}(Y=y, Z=z, X=1)}{\mathbf{P}(X=1)} \\
&= \frac{\mathbf{P}(Y=y, Z=z, X=1)}{\mathbf{P}(X=1)}(\mathbf{P}(X=1)+\mathbf{P}(X=0)) \\
&= \mathbf{P}(Y=y, Z=z, X=1) \\
&\quad + \frac{\mathbf{P}(Y=y, Z=z, X=1)}{\mathbf{P}(X=1)}\,(\mathbf{P}(X=0))
\end{aligned}
$$

为了更清楚的呈现概率的流变，上述推导的最后两步我做了一个不必要的“拆解”。求得新的概率分布如下：

	X	Y	Z	百分比
w_1	1	1	1	$0.116 \times 2 = 0.232$
w_2	1	1	0	$0.274 \times 2 = 0.548$
w_3	1	0	1	$0.009 \times 2 = 0.018$
w_4	1	0	0	$0.101 \times 2 = 0.202$

表 1: $\mathbf{P}(Y, Z \mid X = 1)$ 的联合分布

可以把图（1a）中 (X,Y,Z) 的每一次取值看成构建了一个可能世界，而 **P** 则是给所有 8 个可能世界指派概率。从表（1）中可以看出，条件化 $X=1$，就是把所有 $X\neq 1$ 的世界（即上表（1a）中所有 $X=0$ 的世界）的概率全部褫夺（对应到表 (1) 中就是把所有 $X=0$ 的行都移除），然后再转移到 $X=1$ 的那些世界中去。具体而言，就是根据任意 $X=1$ 的世界在所有 $X=1$ 世界中的概率占比来分配那被褫夺的所有 $X=0$ 的概率。上述的结论可以进一步一般化为：

$$\begin{aligned}\mathbf{P}(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n\mid x_i)&=\frac{\mathbf{P}(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n)}{\mathbf{P}(x_i)}\\&=\frac{\mathbf{P}(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n)}{\mathbf{P}(x_i)}(\mathbf{P}(X_i=x_i)+\sum_{X_i=x,x\neq x_i}\mathbf{P}(X_i\neq x_i))\\&=\mathbf{P}(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n)\\&\quad+\frac{\mathbf{P}(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n)}{\mathbf{P}(x_i)}(\sum_{X_i=x,x\neq x_i}\mathbf{P}(X_i\neq x_i))\end{aligned}$$

上述推导结果说明了在条件化之下，分配谁的概率以及如何分配概率。**首先**，所有满足 $(X_i\neq x_i)$ 世界的概率都要转移给满足 $(X_i=x_i)$ 世界；**其次**，任一 $(X_i=x_i)$ 中的一个世界，比如 $(x_1,\ldots,x_i,\ldots,x_n)$，其所分配到的 $\sum_{X_i=x,x\neq x_i}(X_i\neq x_i)$ 概率的比例，即其在满足 $X_i=x_i$ 的世界集中的概率占比。

3.2 干预

在经典的概率语言中是没有 $\mathbf{P}(Y=y\mid do(X=1))$ 这样的表达式，引入这样表达式的目标是什么呢？首先，要理解因果模型以及与之对应的因果图的意义。所谓因果模型就是对这个世界的一种客观表征，而因果图则定性的描述了数据背后的故事，即对于数据如何生成的一种设想，因果图中暗含着很多定性的假定，比如图（1b）中假定了性别（Z）会影响服药（X），而服药与否则并不影响性别。有了对于客观世界的表征之后，就可以在此基础上拷问出“如果 ... 将会怎样（what if）”这类反事实问题的答案。把两者联系在一起的桥梁正是“干预”（表示为 $do(X=1)$）。从实验的角度，干预做的就是随机对照实验所要实现的目标：去混杂。

其次，do-公式的计算与因果贝叶斯网络相关，所以要理解干预如何让概率流变，要了解因果贝叶斯网络。因果贝叶斯网络涉及一组变量集，一般把它划分为两个子集，内生变量集 **V** 和外生变量集 **U**，根据 **V** 中变量之间直接的因果相干性，可以构建一个与之对应的因果图 **G**（比如，变量 X 对于变量 Y 有直接的因

果影响，就画一个从对应 X 节点到 Y 节点的箭头），且图 **G** 要是一个有向无环图（directed acyclic graph），即从这个图中任意一个点出发，顺着箭头不会最终回到这个点。下图就是一个有向无环图：

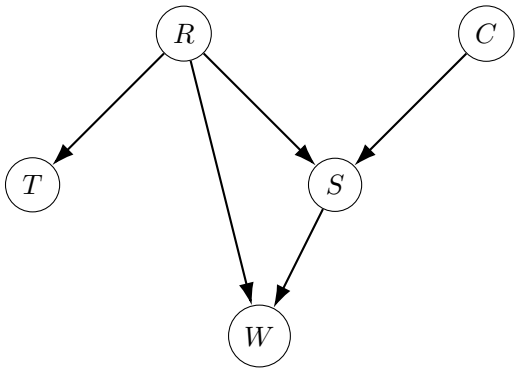
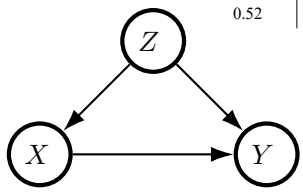


图 2: 一个有向无环图

可以用亲属术语，比如，父、子、子孙、祖先等，来说明有向无环图中节点之间的关系。在上图中，节点 W 的父集为 $\{R, S\}$，W 的祖先集为 $\{R, S, C\}$，S 的子集为 $\{W\}$，C 的子孙集为 $\{S, W\}$。再次，因果贝叶斯网络中存在一个概率分布 **P**，它确切说明了 **U** 中每个变量的概率分布，以及在给定 **V** 中变量的每一个变量的父集情况下，这个变量的概率分布。最后，因果贝叶斯网中的 **P** 和 **G** 满足马尔可夫条件，或者更确切地说，**P** 满足对应于 **G** 的马尔可夫条件：任意变量 $X \in \mathbf{V}$，条件化 **G** 中 X 的父集的情况下 X 的概率独立于 **G** 中 X 的所有非子孙。合并图（1）中的数据表和图得到一个因果贝叶斯网络：

$\mathbf{P}(Z=1)$	$\mathbf{P}(Z=0)$
0.52	0.48



Z	$\mathbf{P}(X=1 \mid Z)$	$\mathbf{P}(X=0 \mid Z)$
1	25/52	27/52
0	25/48	23/48

Z	X	$\mathbf{P}(Y=1 \mid Z,X)$	$\mathbf{P}(Y=0 \mid Z,X)$
1	1	4/5	1/5
1	0	3/5	2/5
0	1	1/3	2/3
0	0	8/23	15/23

图 3: 上述数据和图所对应的因果贝叶斯网络

接下来来考察结构因果分析是如何在干预 $do(X=1)$ 之后来重新进行概率分配的，出于讨论的方便，把变量集 $\{X_1, \ldots, X_n\}$ 分为四个部分：$\{X, Y, PA, K\}$，其中 X 表示干预变量集，Y 表示 X 的子孙集，PA 表示 X 的父集，而 K 则是

X, Y, PA 之外所有其他变量的集合：

$$
\begin{aligned}
&\mathbf{P}(X=x, Y=y, PA=pa, K=k \mid do(X=x)) \\
&\quad= \frac{\mathbf{P}(X=x, Y=y, PA=pa, K=k)}{\mathbf{P}(X=x \mid PA=pa)} && \text{(根据 do-演算)} \\
&\quad= \frac{\mathbf{P}(X=x, Y=y, PA=pa, K=k)}{\mathbf{P}(X=x \mid PA=pa, K=k)} && \text{(根据马尔可夫条件)} \\
&\quad= \frac{\mathbf{P}(X=x, Y=y, PA=pa, K=k)}{\mathbf{P}(X=x, PA=pa, K=k)}\ \mathbf{P}(PA=pa, K=k) \\
&\quad= \frac{\mathbf{P}(X=x, Y=y, PA=pa, K=k)}{\mathbf{P}(X=x, PA=pa, K=k)} \\
&\quad\left(\mathbf{P}\left(PA=pa, K=k, X=x\right)+\sum_{x \neq x} \mathbf{P}\left(PA=pa, K=k, X \neq x\right)\right) \\
&\quad= \mathbf{P}(X=x, Y=y, PA=pa, K=k) \\
&\quad+ \frac{\mathbf{P}(X=x, Y=y, PA=pa, K=k)}{\mathbf{P}(X=x, PA=pa, K=k)}\left(\sum_{x \neq x} \mathbf{P}(PA=pa, K=k, X \neq x)\right)
\end{aligned}
$$

上述 do-演算的推导并不一定需要父变量集，一个满足后门准则的协变量集（a set of covariates）同样也是可以的，此处只是为了讨论的方便。根据关于 $(X = x, Y = y, PA = pa, K = k)$ 对应着一个世界（不妨设为 w）的设定，上述公式可以理解为在干预 $do(X = x)$ 之后，世界 w 的概率分布。在干预之下，概率发生了流变。$\mathbf{P}(X = x, Y = y, PA = pa, K = k)$ 是世界 w 原有概率；$\sum_{X \neq x} \mathbf{P}(PA = pa, K = k, X \neq x)$ 是所有 $(PA = pa, K = k, X \neq x)$-世界的概率和；$\frac{\mathbf{P}(X=x, Y=y, PA=pa, K=k)}{\mathbf{P}(X=x, PA=pa, K=k)}$ 是在 $S_x(PA = pa, K = k, X \neq x)$ 中[2]，世界 w 的概率所占比例。

首先，对于 X 的干预会影响它的子孙 Y 的值，但是 PA 和 K 的值则保持不变，由此也就确定了“最接近世界”：离世界 $w_i = (X = x', Y = y', PA = pa, K = k)$ 最接近的 $(X = x)$-世界是由 $(X = x, PA = pa, K = k)$ 的那些世界所组成的集合（把这个集合记为：$S_x(w_i)$），如下图左图所示：

由于变量之间确切的函数关系未知，在只有数据的条件下，设定 X 的值，并不能唯一确定 Y 的值。每个世界都有可能有不止一个世界和它最接近，多个世界可能有同一个世界集和它们最接近。比如图 1 中，有：$S_{X=1}(w_5) = \{w_1, w_3\}$

[2] $S_x(w)$ 表示的是所有最接近 w 的 $X = x$-世界。

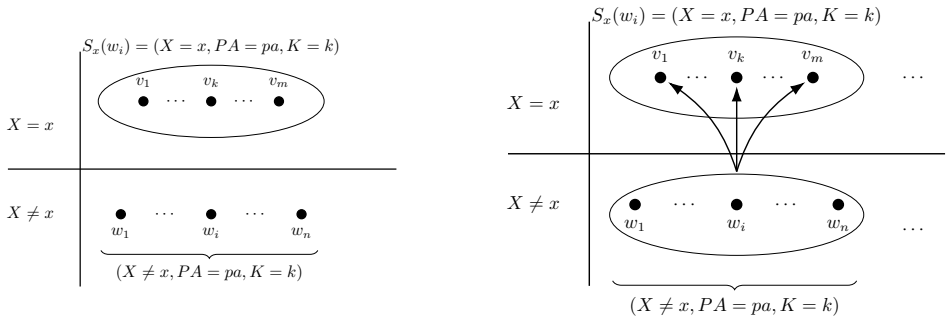


图 4: 左图，世界 $w_1, \ldots, w_n$ 有共同的最接近的 $X = x$-世界集。右图：$\mathbf{P}(w_1)$ 的概率以 $\frac{\mathbf{P}(v_i)}{\mathbf{P}(v_1,\ldots,v_k,\ldots,v_m)}$ 的比例分配到每一个 v_i 世界。

和 $S_{X=1}(w_7) = \{w_1, w_3\}$。于是有了一个判定世界之间相似性的标准：$S_x(w')$ **是由那些** PA, K **的取值和** w' **相同的** $(X = x)$**-世界所组成的集合**。其次，如上面的右图所示，最终得出的公式也说明了每一个 $(X \neq x)$-世界如何把它的概率分配给离它最近的 $(X = x)$-世界中去，根据 $\sum_{X \neq x} \frac{\mathbb{P}(X=x,Y=y,PA=pa,K=k)}{\mathbb{P}(X=x,PA=pa,K=k)}\ \mathbf{P}(PA = pa, K = k, X \neq x)$，分配是成比例的，即每个 w_i 分配给任意一个 v_k 的概率与 $\mathbf{P}(v_k)$ 在 $\mathbf{P}(S_x(w_i))$ 中所占的比例相同。代入计算“服药与康复”的例子，新的概率分布如下表：

	X	Y	Z	百分比
w_1	1	1	1	$0.116 \times \frac{102}{25} = 0.47328$
w_2	1	1	0	$0.274 \times \frac{98}{75} = 0.35802$
w_3	1	0	1	$0.009 \times \frac{102}{25} = 0.03672$
w_4	1	0	0	$0.101 \times \frac{98}{75} = 0.1320$

表 2: $\mathbf{P}(Y, Z \mid do(X = 1))$ 的联合分布

表 1 和表 2 生动的说明了对两种条件句不同数学刻画：“如果观察到服药的情况下 $(X = 1)$，那么 $\cdots$”和“如果干预使服药的情况下 $(do(X = 1)$，那么 $\cdots$”。前者只要使用条件概率就可求的，后者则需要借助因果图的帮助以求得最终的干预条件句的概率。前者刻画的不必然是因果关系，也许是一种相关关系。后者则是基于因果图或者因果机制的推断，其所得到的将是前件事实对于后件事实的因果影响。

3.3 影像

有别于条件概率，大卫·刘易斯设想出一种在可能世界的框架内给反事实条件句指派概率的理论：影像。（[12]）有必要介绍一下刘易斯关于可能世界本体论预设。首先，刘易斯预设可能世界是有穷的，用 w 表示一个可能世界，且每个世界 w 都有一个概率 $\mathbf{P}(w)$，所有世界的概率加起来总和为 1。可能世界的有穷假定对当前的讨论而言很合理，因为如果只是做纯粹的理论探讨，可以设想无穷可数的可能世界，但是在讨论具体反事实的真值条件和概率的时候，正如表（1a）所体现的那样，实际所考虑的世界远没有那么庞大。用 w_A 表示一个最接近 w 的 A-世界（$S_A(w)$ 表示最接近 w 的一个 A-世界集），$[\![A]\!]$ 表示所有 A 在其中为真的世界的集合。假定如果 A 是可能的，表示 A 在世界 w 中为真如下：

$$w(A) =^{df} \begin{cases} 1 & \text{如果 } w \in [\![A]\!] \\ 0 & \text{否则} \end{cases}$$

其次，这个影像理论是以斯托奈克（[19]）的语义理论为基础。尽管正如刘易斯（[10]）所批评的，斯托奈克（[19]）的语义的**极限假定**（Limit Assumption）和**反对称**（Anti-symmetry）限制（这个限制禁止了“**结**（ties）”的存在，使得世界之间是一种全序关系）有诸多反例。但是这个语义却和因果模型的语义有很多共通之处，前者通过选择函数 $f(w, A)$ 确定唯一最接近 w 的 A-世界，而后者通过干预也确定了所有变量的一次取值，而干预所改变的结构方程也是与其干预变量对应的方程，其他方程则保持不变。

注意，刘易斯的最终语义是不预设极限假定的（[11]），存在无限接近，但是又没有最接近的可能世界的情形，尽管在理论上有更强的包容性，但是在现实的反事实判定中则不必要。现实中在思考一些反事实真值的时候的确也不会那么精细化（fine-grained），身高一米七五的我确信这样一个反事实：“如果我身高高于一米八，我的命运将会不同”的时候，我并不是在刘易斯的意义上认为存在无限接近但是又没有最接近的身高高于一米八的世界，然后再确定一个身高高于一米八-我命运不同的世界比任何身高高于一米八-我命运没有不同的世界更接近现实的世界，于是反事实为真。根据斯塔内克的语义：

$$w(A \mathrel{\square\!\!\rightarrow} C) = w_A(C)，\text{如果 } A \text{ 是可能的} \tag{9}$$

再次，设定任意句子 A 的概率是它在其中为真的世界的概率和：

$$\mathbf{P}(A) = \sum_w \mathbf{P}(w) \cdot w(A) \tag{10}$$

有了上述概念准备，就能够定义对概率 **P** 加诸 A 的影像了（表示为 $\mathbf{P}^A$）。在给出具体的形式定义之前得要把这个操作要实现的目标讲清楚。首先，这个操作要实现把每个世界 w' 的概率转移到 w'_A 中去，即最接近 w' 的 A-世界中去了。但是世界 w' 中的概率只是被转移，没有“消失”，所以剩下所有 w'_A 世界的概率和加起来还是 1；其次，每一个 A 世界保有它原先的概率，因为如果 w' 是 A-世界，那么 $w'_A = w'$，且它还能获得额外的，如果存在 $\neg A$-世界 w 且 $w_A = w'$（即如果最接近 w 的 A-世界是 w'），那么 w 的概率就会转移到 w' 中去；再次，每一个 $\neg A$-世界则不能保有它初始的概率，也不会获得额外的概率，因为最接近它的 A-世界必然不是自身。所以对概率 **P** 加诸 A 的影像之后，概率质量就都集中到了 A-世界中。正如刘易斯所说，对 **P** 加诸 A 的影像的目的是为了实现确保 A 成立情况之下的**极小修正**：

> 对 **P** 加诸 A 的影像给予了在如下意义上的极小修正（*minimal revision*）：与其他修正 **P** 以使得 A 成立的径路不同，在它（即“对 **P** 加诸 A 的影像”）中不会发生一个世界会无端地把它的概率转移给那些与它不相似的世界。（[12]，第 311 页）

可以把上述的想法图像化如下（5）：

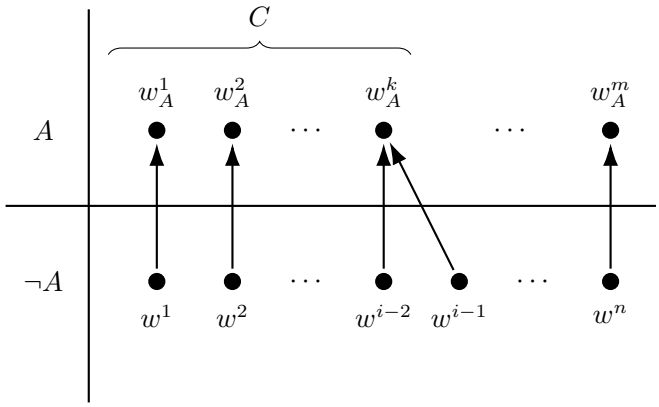


图 5: Lewis [12] 影像理论的直觉图

如上图所示，可能存在一个 A-世界，它不是任何一个 $\neg A$-世界的最接近世界，但是任何一个 $\neg A$-世界必然存在一个最接近它的 A-世界，只要 A 是可能的。[3]也可能存在一个 A-世界，它同时是多个 $\neg A$-世界的最接近世界。于是对 **P** 加诸 A 的

[3]这是斯塔内克语义的一个重要预设：任意世界 w 和命题 A，如果存在可能世界 A 在其中为真，那么存在唯一最接近 w 的 A-世界：w_A。

影像可以形式定义如下：

$$\mathbf{P}^{A}(w)=\sum_{w'}\mathbf{P}(w')\cdot\mathbf{P}(w\mid w'_{A}) \tag{11}$$

上式本质上就是逐个考察可能世界 w'，如果 w 是最接近 w' 的 A-世界，那么我们就把 w' 的概率转移给可能世界 w。这里的符号表达可能有点不太严格，主要也是为了后面讨论的方便（如果把命题理解为它在其中为真的所有可能世界的集合）：

$$\mathbf{P}(w\mid w'_{A})=\frac{\mathbf{P}\left(\{w\}\bigcap\{w'_{A}\}\right)}{\mathbf{P}\left(\{w'_{A}\}\right)}=\begin{cases}1 & \text{如果 } w\text{是最接近 } w'\text{的}A-\text{世界}\\ 0 & \text{否则}\end{cases}$$

有了式（11）就可以进一步计算在 $\mathbf{P}^{A}$ 之下，C 在其中的概率（[12]，第 311 页）就可以计算如下：

$$\begin{aligned}
\mathbf{P}^{A}(C)&=\sum_{w'}\mathbf{P}^{A}(w')\cdot w'(C) && \text{由 (10)}\\
&=\sum_{w'}\left(\sum_{w}\mathbf{P}(w)\cdot\mathbf{P}(w'\mid w_{A})\right)\cdot w'(C) && \text{由 (11)}\\
&=\sum_{w}\mathbf{P}(w)\left(\sum_{w'}\mathbf{P}(w'\mid w_{A})\cdot w'(C)\right) && \text{由代数法则}\\
&=\sum_{w}\mathbf{P}(w)\cdot w_{A}(C) && \text{最接近 } w \text{ 的 } A\text{-世界只有一个}\\
&=\sum_{w}\mathbf{P}(w)\cdot w(A\ \square\!\!\rightarrow C) && \text{由 (9)}\\
&=\mathbf{P}(A\ \square\!\!\rightarrow C) && \text{由 (10)}
\end{aligned} \tag{12}$$

即反事实条件句 $A\ \square\!\!\rightarrow C$ 的概率 $\mathbf{P}(A\ \square\!\!\rightarrow C)$ 就是对概率 $\mathbf{P}$ 加诸 A 的影像之后所得到概率分布中 C 的概率。[4]

[4]上述刘易斯的证明非常优美，但是对于理解概率的流变，我给出如下一个更符合直觉的证明：

证明. 如果 $w\in[\![A]\!]$ 且 $w(A\ \square\!\!\rightarrow C)=1$，那么根据方程（9）有 $w_{A}(C)=1$，根据斯塔内克的语义，有 $w_{A}(A\ \square\!\!\rightarrow C)=1$。

如果 $w\in[\![A]\!]$ 且 $w(A\ \square\!\!\rightarrow C)=0$，那么根据方程（9）有 $w_{A}(C)=0$，根据 Centering，有 $w_{A}(A\ \square\!\!\rightarrow C)=0$ 对 $\mathbf{P}$ 做关于 A 的影像

1. 如果 $w\notin[\![A]\!]$ 且 $w(A\ \square\!\!\rightarrow C)=1$，$w$ 的概率会转移到 w_{A} 世界中去，且如上已证 $A\ \square\!\!\rightarrow C$ 在 w_{A} 世界中也为真。所以尽管世界 w 的概率被褫夺，但它只是转移到了另一个 $A\ \square\!\!\rightarrow C$ 在其中为真的世界。所以得出，影像之后 $A\ \square\!\!\rightarrow C$ 在其中为真的所有那些世界的概率和并没有减少。

接下来的问题，可以把这样一个在可能世界框架中的影像理论应用于图（1）所述情形的分析吗？即在上述数据和因果图的前提下，可以在影像框架内考察反事实 $(X = 1) \boxright (Y = 1)$ 吗？答案是否定的。影像理论并没有给出一个标准以确定最接近 w 的 A-世界。所以影像并不具有现实可操作性，尽管思想优美且建构精巧。

也可以给出强行判定相似性的标准。比如可以给出如下素朴的标准：尽可能多的事实的符合。显然这并不是一个好的标准，有很多反例（[20]），甚至连刘易斯也承认有很多细节之处值得深究："是否具体事实的近似相似性应当有很小的权重或者没有权重这个问题是一个很好的问题。不同的情形所得出的结果是不同的，我想知道其中的原因。"（[13]，第 472 页）一般称这种理解为对于相似性概念的"直觉式"理解。并以此来判定最接近的 $X = 1$-世界。那么 $(X = 0, Y = 1, Z = 1)$ 的最接近的 $(X = 1)$-世界就是 $(X = 1, Y = 1, Z = 1)$，同理可以确定剩下 $(X = 0)$-世界所最接近的 $(X = 1)$-世界，按照影像，把概率转移到这些最接近的 $(X = 1)$-世界，有如下结果：

X	Y	Z	百分比
1	1	1	0.116 + 0.334= 0.45
1	1	0	0.274 + 0.079 = 0.353
1	0	1	0.009 + 0.051= 0.06
1	0	0	0.101 + 0.036 = 0.137

表 3: 对 **P** 加诸 $X = 1$ 的影像之后的概率分布

于是：

$$\mathbf{P}^{(X=1)}(Y = 1) = 0.45 + 0.353 = 0.803$$

尽管最终的结果和干预演算所算出来的结果（$0.47328 + 0.35802 = 0.8313$）差不多，但实质上并不具有合理性。最大的问题在于，可能世界框架关于反事实条件句概率判定的设想还是从对于那些日常反事实的概率赋予的直觉发展出来的，世界 w 和 w_A 在直觉中是清晰明确的，但这种清晰明确是有限度的，而干预演算则给出了一体的方案来确定最接近的世界，以及在干预之下来如何来分配概率。

2. 如果 $w \notin [\![A]\!]$ 且 $w(A \boxright C) = 0$，w 的概率会转移到 w_A 世界中去，且如上已证 $A \boxright C$ 在 w_A 世界中也不为真。所以 $A \boxright C$ 在其中并不为真的世界 w 的概率被转移到了另一个 $A \boxright C$ 在其中并不为真的世界。所以得出，影像之后 $A \boxright C$ 在其中为真的所有那些世界的概率和并没有增加。

□

4 修正影像理论及其蕴涵

接下来将在刘易斯原有框架的基础上，参考干预的思路，给出世界之间相似性的判定标准，并设定 $\neg A$ 世界的概率如何在最接近它的 A-世界中分配。令 A 为 $X = x'$，w 为 $(X = x, Y = y, PA = pa, K = k)$。如果 $x \neq x'$，那么如果对 **P** 加诸 A 的影像，w 的概率将会被褫夺，所以 $\mathbf{P}^A(w) = 0$；如果 $x = x'$，如果对 **P** 加诸 A 的影像，那么 w 不仅能保有原先的概率，而且能分得所有 $\sum_{X\neq x}(X \neq x, PA = pa, K = k)$ 世界的部分概率，不妨设 $\forall w' \in (X \neq x, PA = pa, K = k)$，显然 $S_x(w') = (X = x, PA = pa, K = k)$。在这个思路之下来计算 $\mathbf{P}^A(w)$：

$$
\begin{aligned}
\mathbf{P}^A(w) &= \sum_{w'\in(X\neq x, PA=pa, K=k)} \big(\mathbf{P}(w) + \mathbf{P}(w \mid S_x(w'))\mathbf{P}(w')\big) \\
&= \mathbf{P}(w) + \sum_{w'\in(X\neq x, PA=pa, K=k)} \mathbf{P}(w \mid S_x(w'))\mathbf{P}(w') \\
&= \mathbf{P}(w)\mathbf{P}(w \mid S_x(w')) + \sum_{w'\in(X\neq x, PA=pa, K=k)} \mathbf{P}(w \mid S_x(w'))\mathbf{P}(w') \\
&\qquad\qquad \text{(因为 } S_x(w) = w\text{)} \\
&= \sum_{w'\in(X\neq x, PA=pa, K=k)\cup\{w\}} \mathbf{P}(w')\mathbf{P}(w \mid S_x(w')) \\
&= \sum_{w'\in(X\neq x, PA=pa, K=k)\cup\{w\}} \mathbf{P}(w')\mathbf{P}(w \mid S_x(w')) \\
&= \sum_{w'\in(X\neq x)\cup\{w\}} \mathbf{P}(w')\mathbf{P}(w \mid S_x(w')) \\
&\qquad\qquad \text{(如果 } w' \in (X \neq x)/(X \neq x, PA = pa, K = k)\text{，} \\
&\qquad\qquad\qquad \text{那么 } w \notin S_x(w')\text{，则 } \mathbf{P}(w \mid S_x(w'))\text{=0)} \\
&= \sum_{w'} \mathbf{P}(w')\mathbf{P}(w \mid S_x(w')) \\
&\qquad \text{(因为如果 } w' \in (X = x)/w\text{，那么 } w' = S_x(w')\text{，于是 } \mathbf{P}(w \mid w') = 0\text{)}
\end{aligned}
$$

于是最终得到**修正影像理论**。本文所得出的最终结论和珀尔文章的结果相同，但是证明的径路有差异（[16]，第 3 页），珀尔是直接给出了两个规定得出了结果，我则是在 do-演算的公式推演中试图找到概率流变分配的规律，并以此出发来得出 (13)：

$$
\mathbf{P}^A(w) = \sum_{w'} \mathbf{P}(w')\mathbf{P}(w \mid S_A(w')) \tag{13}
$$

由于修正影像理论完全是按照干预的思路来建构的，所以可以证明 $\mathbf{P}(w \mid$

$do(A)) = \mathbf{P}^A(w)$。证明的思路和上述建构思路基本一致。同样地，式 (13) 中的条件概率意义如下：

$$\mathbf{P}(w \mid S_x(w')) = \frac{\mathbf{P}\left(\{w\} \bigcap S_x(w')\right)}{\mathbf{P}\left(S_x(w')\right)} = \begin{cases} 0 & \text{如果 } w \notin S_x(w') \\ \dfrac{\mathbf{P}(w)}{\mathbf{P}(S_x(w'))} & \text{如果 } w \in S_x(w') \end{cases}$$

首先，修正影像理论实质已经重新设定了刘易斯影像理论的基本预设，唯一性假定不再成立，任意一个 $(X \neq x)$-世界 w_i，存在不止一个 $(X = x)$-世界同等最接近它，于是斯塔内克语义不再适用，即公式（9）不再成立。需要给出新的语义，根据可能世界框架下的反事实语义讨论，在不考虑那种无限接近但没有最接近的情形下，即预设极限假定，很自然的采用刘易斯的语义（[11]，第 422 页），形式化的表达如下：

$$w(A \mathrel{\square\!\!\rightarrow} C) = \begin{cases} 1 & \text{if} \quad S_A(w) \subseteq [\![C]\!] \\ 0 & \text{if} \quad S_A(w) \nsubseteq [\![C]\!] \end{cases} \tag{14}$$

上式表达的是，如果反事实 $A \mathrel{\square\!\!\rightarrow} C$ 在 w 中为真，那么 C 在所有最接近 w 的 A-世界中为真。其中 $[\![C]\!]$ 表示 C 在其中为真的所有可能世界的集合。有了方程（13），立刻可以得出：

$$\mathbf{P}^A(C) = \sum_{w \in [\![C]\!]} \sum_{w'} \mathbf{P}(w')\mathbf{P}(w \mid S_A(w')) \tag{15}$$

有了（13）和（15），就可以讨论服药与康复的例子：

$$\begin{aligned} \mathbf{P}^{X=1}(w_1) &= \sum_{w'} \mathbf{P}(w')\mathbf{P}(w_1 \mid S_{X=1}(w')) \\ &= \mathbf{P}(w_5)\mathbf{P}(w_1 \mid S_{X=1}(w_5)) + \mathbf{P}(w_7)\mathbf{P}(w_1 \mid S_{X=1}(w_7)) + \mathbf{P}(w_1) \\ &= \mathbf{P}(w_5) \cdot \frac{\mathbf{P}(w_1)}{\mathbf{P}(w_1) + \mathbf{P}(w_3)} + \mathbf{P}(w_7) \cdot \frac{\mathbf{P}(w_1)}{\mathbf{P}(w_1) + \mathbf{P}(w_3)} + \mathbf{P}(w_1) \\ &= 0.47328 \end{aligned} \tag{16}$$

同理可以求出表 2 其他的值，与方程计算的结果相同。

其次，基于可能世界框架的修正影像理论隐含地贯彻了一种使真者语义（[5]）的精神，使得关于析取前件的反事实处理问题成为可能，而且在形式上要比筛选语义在形式上更为简洁和优美。（[18]）而且正如珀尔所证明的（[16]），它还能够处理析取前件反事实。

需要做一点背景介绍。相较于反事实的可能世界逻辑（[10]），结构反事实逻辑的一个短板在于它的语言只能涵盖部分反事实，而不像前者一样理论上可以处理任意复杂的反事实条件句。扩充结构反事实逻辑语言的努力一直在进行中。一开始只能给相当窄的一类反事实赋予真值：$(A_1 \wedge \ldots \wedge A_n) \boxminus\!\!\rightarrow (C_1 \wedge \ldots \wedge C_n)$；（[6]）后来能够处理后件是布尔组合，前件是事件合取的条件句赋予真值（[7, 8]）$(A_1 \wedge \ldots \wedge A_n) \boxminus\!\!\rightarrow C$；目前最优的情形是能够判定任意复杂布尔组合前件和后件的条件句（[2]），但是不能赋予如下反事实以真值：$A \boxminus\!\!\rightarrow ((B \boxminus\!\!\rightarrow C) \boxminus\!\!\rightarrow D)$，因为后件中的嵌套反事实没有一个布尔命题作为前件。

尽管在逻辑上已经扩展了结构反事实的语言使其能涵盖析取行动的反事实，但是如何把这种逻辑上的成果转化到实际的干预演算的计算中去，至今依然是一个悬而未解的问题，问题的根源在于干预演算并没有清楚的设定它是如何把 $\neg A$-世界的概率分配到 A-世界的，如图（6）所示：

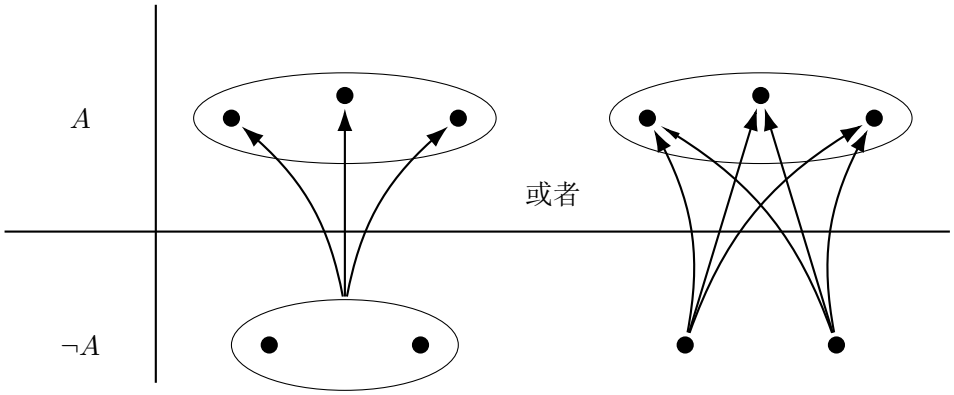


图 6: 干预演算的概率分配示意图

公式 $\frac{\mathbf{P}(X=x,Y=y,PA=pa,K=k)}{\mathbf{P}(X=x,PA=pa,K=k)} \sum_{X\neq x} \mathbf{P}(PA = pa, K = k, X \neq x)$ 并没有清楚地说明分配方式：是把 $\neg A$-世界中那些有共同最接近世界的概率加和在一起，然后再按比例分配给那些最接近 A-世界，还是每一个 $\neg A$-世界都单独按比例分配它的概率给最接近它的世界集？在数学上，怎么来分配并没有问题，因为这两种分配方式是等价的，但是在给析取前件反事实指派概率时，分配方式就起决定性作用了。

考察如下图一个简单的例子（7），其中变量边上括号中的内容表示的是变量的可能取值，比如 X 的旁边的 (x_1, x_2) 表示 X 的两个可能取值 x_1 和 x_2，所以有了类似于（1a）那样的数据表，要计算 $\mathbf{P}(d_1 \mid do(x_1) \text{ or } do(y_1))$，但是因果结构模型的框架并不能处理类似这样的析取前件反事实。

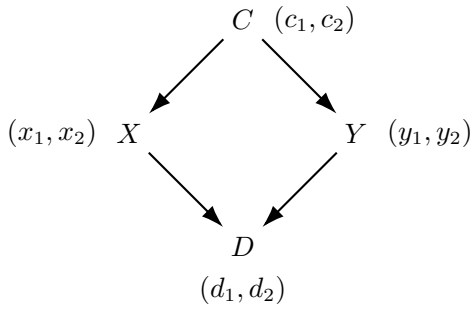


图 7: 一个因果图及所对应的数据

修正影像理论（正如在方程 (16) 的运算中所体现出来的那样）则绝对按照如下方式来分配概率的：每一个 $\neg A$-世界都单独地（而不是总体加和）来分配概率。如图（6）的右图所示。所以尽管不能通过 do-演算（[15]，第 85 页）计算 $\mathbf{P}(d_1 \mid do(x_1) \text{ or } do(y_1))$，但是却可以计算 $\mathbf{P}^{(x_1 \vee y_1)}(d_1)$，根据概率流变的思路具体计算如下：

$$\begin{aligned}\mathbf{P}^{x_1 \vee y_1}(d, c, x, y) &= \mathbf{P}(d, c, x, y) + \frac{\mathbf{P}(d, c, x, y)}{\mathbf{P}(x_1 \vee y_1, c)} \mathbf{P}(x_2, y_2, c) \\ &= \mathbf{P}(d, c, x, y) \frac{\mathbf{P}(x_1 \vee y_1, c) + \mathbf{P}(x_2, y_2, c)}{\mathbf{P}(x_1 \vee y_1, c)} \\ &= \frac{\mathbf{P}(d, c, x, y)}{\mathbf{P}(x_1 \vee y_1 \mid c)}\end{aligned}$$

其中 x, y 必有一个取值为 x_1 或者 y_1，进一步得到：

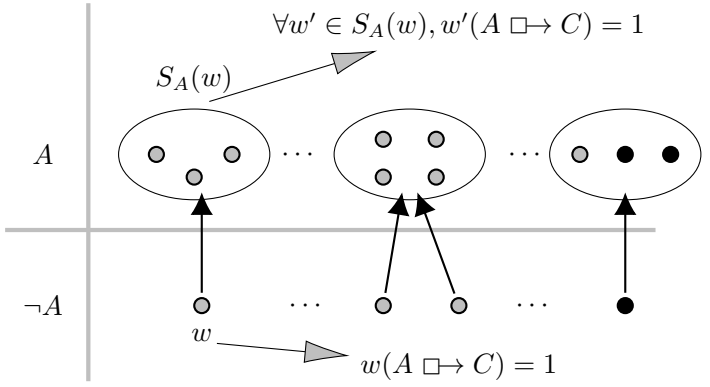
$$\mathbf{P}^{(x_1 \vee y_1)}(d_1) = \sum_c \frac{\mathbf{P}(d_1, x_1 \vee y_1, c)}{\mathbf{P}(x_1 \vee y_1 \mid c)}$$

再次，扩展的影像理论是在新的反事实条件句的语义（14）之下被引入的，在原先语义（9）的基础上，可以证明公式（12），但是在新的语义基础上，并不能证明类似公式（12）的结论。事实上，可以形式的证明 $\mathbf{P}^A(C)$ 和 $\mathbf{P}(A \boxdot\!\!\rightarrow C)$ 不只是不相等，而且 $\mathbf{P}^A(C) \geq \mathbf{P}(A \boxdot\!\!\rightarrow C)$：

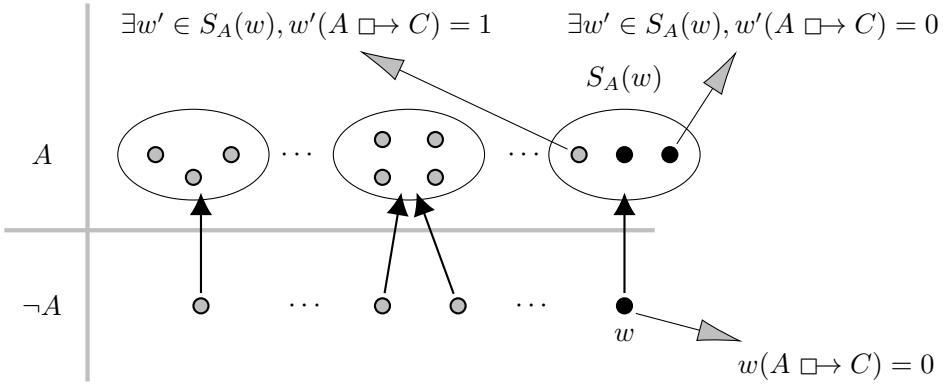
证明. 如果 $w \notin [\![A]\!]$ 且 $w(A \boxdot\!\!\rightarrow C) = 1$，那么根据方程（14）有 $S_A(w) \subseteq [\![C]\!]$，根据 Centering，$\forall w' \in S_A(w), w'(A \boxdot\!\!\rightarrow C) = 1$。

如果 $w \notin [\![A]\!]$ 且 $w(A \boxdot\!\!\rightarrow C) = 0$，那么根据方程（14）有 $S_A(w) \nsubseteq [\![C]\!]$，根据新的语义设定，可能存在 $w' \in S_A(w), w'(A \boxdot\!\!\rightarrow C) = 1$。

对 **P** 做关于 A 的影像。首先，如果 $w \notin [\![A]\!]$ 且 $w(A \boxdot\!\!\rightarrow C) = 1$，$w$ 的概率会转移到 $S_A(w)$ 世界中去，且如上已证 $S_A(w)$ 中任一世界都是一个 $A \boxdot\!\!\rightarrow C$ 在其中为真的世界。所以尽管世界 w 的概率被褫夺，但它只是转移到 $A \boxdot\!\!\rightarrow C$ 在其中为真的一个世界集 $S_A(w)$ 中去了。所以得出，影像之后，$A \boxdot\!\!\rightarrow C$ 在其中为真的所有那些世界的概率和并没有减少。



其次，如果 $w \notin [\![A]\!]$ 且 $w(A \boxdot\!\!\rightarrow C) = 0$，$w$ 的概率会转移到 $S_A(w)$ 世界中去，且如上已证，以及上述例子所显示的那样，$S_A(w)$ 中存在一个世界 w'，$A \boxdot\!\!\rightarrow C$ 在其中为真。但是 w' 世界却分享了 $A \boxdot\!\!\rightarrow C$ 不在其中为真的世界 w 的概率，所以得出，影像之后，$A \boxdot\!\!\rightarrow C$ 在其中为真的所有那些世界的概率和会增加。于是有 $\mathbf{P}^A(C) \geq \mathbf{P}(A \boxdot\!\!\rightarrow C)$。



□

原本期待的类似式 (12) 的结果的证明过程如下：

$$\begin{aligned}
\mathbf{P}^{A}(C) &= \sum_{w\in[\![C]\!]}\sum_{w'}\mathbf{P}(w')\,\mathbf{P}(w \mid S_A(w')) \\
&= \sum_{w'}\mathbf{P}(w')\sum_{w\in[\![C]\!]}\mathbf{P}(w \mid S_A(w')) \\
&= \sum_{w'}\mathbf{P}(w')\,\mathbf{P}(C \mid S_A(w')) \qquad (17)\\
&= \sum_{w'}\mathbf{P}(w')\,w'(A \boxdot\!\!\rightarrow C) \qquad (18)\\
&= \mathbf{P}(A \boxdot\!\!\rightarrow C)
\end{aligned}$$

但因为从（17）到（18）并不总是成立的，只有在如下条件下才能够相互替换：

$\mathbf{P}(C \mid S_A(w')) = 1$ 当且仅当 $S_A(w') \subseteq [\![C]\!]$ 当且仅当 $w'(A \boxdot\!\!\rightarrow C) = 1$

即只有在图（8）所示的情况下才是成立的。

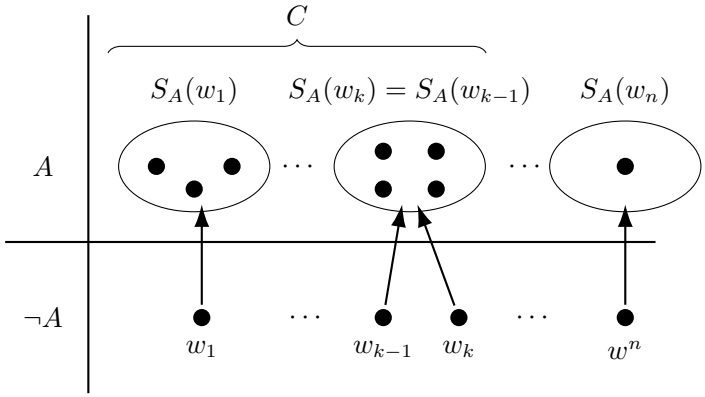


图 8: 原本所设想的修正影像理论的直觉图

但是正如计算服药与康复例子中的 $\mathbf{P}^{X=1}(Y=1)$，有 $\mathbf{P}^{X=1}(Y=1) = \mathbf{P}^{X=1}(w_1) + \mathbf{P}^{X=1}(w_2)$，$w_1$ 和 w_2 来自两个不同的最接近的 $X=0$ 世界，即 $S_{X=1}(w_5) \nsubseteq (Y=1)$ 。

5 结语

本文讨论了三种改变信念度的方式，条件化 $\mathbf{P}(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n \mid x_i)$，干预（$\mathbf{P}(x,y,pa,k \mid do(x))$）和影像（$\mathbf{P}^{X=x}(w)$）。可以通过如下两个问题来对它们的概率流变方式作一个说明：第一、谁的概率被夺走以及分配给谁，第二、具体具体的分配方案为何。在可能世界的框架内，上述两个问题可以重述为：第一、判定世界之间接近性的标准是什么？第二、如果最近世界不止一个，那么按照什么分配方式，把褫夺的概率分配给那些最接近的世界？上述三种方式都给出了（或者部分给出了）自己的解决方案。

条件化的思路是先把所有 $\sum_{X\neq x_i}(X \neq x_i)$-世界的概率全部夺走并分配给所有 $\sum_{X=x_i}(X = x_i)$-世界；再分配给任一 $(X = x_i)$-世界的概率比例就是这个 $(X = x_i)$-世界在所有 $\sum_{X=x_i}(X = x_i)$-世界中的概率占比。**干预**的思路是，先全部世界被划分为形如 $(PA = pa, K = k)$ 的子类，任一子类 $(PA = pa, K = k)$ 又被划分为两个子类：$(PA = pa, K = k, X = x)$ 和 $(PA = pa, K = k, X \neq x)$，干预（$do(x)$）之后，所有 $\sum_{X\neq x}(PA = pa, K = k, X \neq x)$-世界的概率被夺走，分配给所有 $\sum_{X=x}(PA = pa, K = k, X = x)$-世界。所以，与任一 $(X \neq x)$-世界最接近的世界，是那些与它有共同 $PA = pa$ 和 $K = k$ 的 $(X = x)$-世界。再分配给任一 $(PA = pa, K = k, X = x)$-世界的概率比例就是这个 $(PA = pa, K = k, X = x)$-世界在所有 $\sum_{X=x}(PA = pa, K = k, X = x)$-世界中的概率占比。条件化和干预概率分配的模式其实是一样的，不同之处在于确定最接近世界的方式，后者要更细化，在父集 PA 和 K 集取值相同的世界之间才能进行概率的传递。**影像**的思路是先任一 $(X \neq x)$-世界的概率会被夺走。再把这个世界的概率分配给与它最接近的那个 $(X = x)$-世界。相较于前两者而言，影像理论最单薄，既没有判定世界相似性的标准，还有一个非常强的设定，即对于每一个 $(X \neq x)$-世界，都存在唯一一个与它最接近的 $X = x$-世界。

从干预的视角来看，这个设定太强了，不防松动一下，设定存在多个 $(X = x)$-世界，同等接近于一个 $(X \neq x)$-世界。但是这个更为宽松的设定也引入了新的问题：如果存在多个最接近任一 $X \neq x$-世界的世界，该如何把这个 $X \neq x$-世界的概率分配给那些最接近于它的 $X = x$-世界们？我们所做的是用干预如何确定最接近的世界，以及如何分配概率的思路来修正刘易斯的影像理论。其实完全可以不考虑如何来确定最接近的世界，只设定如果一个世界有多个最接近于他的世界，那么这个世界要把它的概率分配给所有这些最接近它的世界，以及任一最接近世界分配到的比例是这个世界在所有最接近世界中的概率占比。但是这个很直观的推广却让刘易斯原本的结论（式 (12)）不再成立，尽管修正影像理论拥有原本的

干预理论所没有的新特征：处理析取前件反事实。

最后，还有如下一些本文付之阙如但有待进一步深入的问题。首先，如果要让修正影像理论能够得到刘易斯的原本结论，如何来合理地设定概率的分配方式以实现目标；其次，正如刘易斯所宣称的一样，影像理论毕竟是一种给全部反事实条件句指派概率的理论，根据珀尔的因果三阶梯（[17]），干预演算只是处理了第二个层级的反事实，即展望式反事实（prospective counterfactuals），而不是反省式反事实（retrospective counterfactuals），而后者的计算步骤和前者是不一样的，它的概率流变会如何，会得到影像理论一样的结论吗？当然，这个理论需要的不只是因果贝叶斯网络，而且需要确切的结构方程，且给人的感觉，它的概率流变和我们所讨论这些已经很不相同，它会有什么有趣的蕴涵和意义，也有待进一步的辛劳。[5]

参考文献

[1] R. Bradley, 2012, “Multidimensional possible-world semantics for conditionals”, *Philosophical Review*, **121(4):** 539–571.

[2] R. Briggs, 2012, “Interventionist counterfactuals”, *Philosophical studies*, **160(1):** 139–166.

[3] K. Fine, 1975, “Review of Lewis’ counterfactuals”, *Mind*, **84(335):** 451–458.

[4] K. Fine, 2012, “Counterfactuals without possible worlds”, *The Journal of Philosophy*, **109(3):** 221–246.

[5] K. Fine, 2017, “Truthmaker semantics”, *A Companion to the Philosophy of Language*, **Vol. 2**, pp. 556–577, Wiley Online Library.

[6] D. Galles and J. Pearl, 1998, “An axiomatic characterization of causal counterfactuals”, *Foundations of Science*, **3(1):** 151–182.

[7] J. Y. Halpern, 2000, “Axiomatizing causal reasoning”, *Journal of Artificial Intelligence Research*, **12**: 317–337.

[8] E. Hiddleston, 2005, “A causal theory of counterfactuals”, *Noûs*, **39(4):** 632–657.

[9] C. Hitchcock, 2001, “The intransitivity of causation revealed in equations and graphs”, *The Journal of Philosophy*, **98(6):** 273–299.

[10] D. Lewis, 1973, *Counterfactuals*, Cambridge, MA: Harvard University Press.

[11] D. Lewis, 1973, “Counterfactuals and comparative possibility”, *Journal of Philosophical Logic*, **2(4):** 418–446.

[12] D. Lewis, 1976, “Probabilities of conditionals and conditional probabilities”, *The Philosophical Review*, **85(3):** 297–315.

[5]谢谢香港中文大学的张寄冀老师，以及中山大学的方静芝博士在讨论这个问题时给予我的巨大帮助。

[13] D. Lewis, 1979, “Counterfactual dependence and time’s arrow”, *Noûs*, 455–476.

[14] J. Pearl, 2000, *Causality: Models, Reasoning, and Inference.* First Edition, New York: Cambridge University Press.

[15] J. Pearl, 2009, *Causality: Models, Reasoning, and Inference*, Second Edition, New York: Cambridge university press.

[16] J. Pearl, 2017, “Physical and metaphysical counterfactuals: Evaluating disjunctive actions”, *Journal of Causal Inference*, **5(2):** 1–10.

[17] J. Pearl and D. Mackenzie, 2018, *The Book of Why: The New Science of Cause and Effect*, Basic Books.

[18] P. Santorio, 2019, “Interventions in premise semantics”, *Philosophers’ Imprint*, **19(1):** 1–27.

[19] R. Stalnaker, 1968, “A theory of conditionals”, *IFS: Conditionals, Belief, Decision, Chance and Time*, pp. 41–55, Dordrecht: Springer Science & Business.

[20] P. Tichý, 1976, “A counterexample to the Stalnaker-Lewis analysis of counterfactuals”, *Philosophical Studies*, **29(4):** 271–273.

[21] J. Woodward, 2016, “Causation in science”, in P. Humphreys(ed.), *The Oxford Handbook of Philosophy of Science*, pp. 163–183, New York: Oxford University Press.

（责任编辑：执子）

Conditionalization, Intervention and Imaging — Three Ways of Changing the Degree of Belief

Xiao'an Wu

Abstract

This paper discusses three ways in which degree of beliefs can change. The first two are ways of how a perfectly rational agent can change the degree of belief given new observational or intervening information; the latter is an imaging theory of how probabilities are assigned to counterfactuals within the framework of possible worlds, and likewise a theory of how probabilities flow. I will illustrate and compare how the above three changes in degrees of belief can be understood within the framework of possible worlds. And Judea Pearl modify Lewis's imaging theory with reference to the workings of the intervention theory, and this revised imaging theory has advantages over the intervention theory, such as the treatment of analytic antecedent counterfactuals, but because this revision replaces the Robert Stalnaker semantics that Lewis originally adopted with his own, I will show that it will no longer yield the original result that $\mathbf{P}^A(C) = \mathbf{P}(A \mathrel{\square\!\!\rightarrow} C)$.

Xiao'an Wu　School of Marxism, Northwestern Polytechnical University
wuxiaoan1984@126.com