arXiv:2602.16612v1 [cs.LO] 18 Feb 2026

因果与组合抽象

RobinLorenz SeanTull{
robin.lorenz,sean.tull}@quantinuum.com
Quantinuum，英国伦敦

抽象

从低层级抽象到 *more explanatory high level* 描述层级，并且理想情况下以符合变量各自 *causal structures* 的方式进行，可以说是科学实践、因果推断问题以及稳健、高效、可解释的人工智能的基础。我们将低层级模型与高层级模型之间的抽象作为 *natural transformations*进行一般性阐述，并重点讨论因果模型。这为 *causal abstraction*提供了一种新的形式化方法，统一了文献中的多种概念，包括构造性因果抽象、Q-τ 一致性、基于交换干预的抽象，以及“分布式”因果抽象。我们的方法以 *categorytheory*的形式进行形式化，并采用一种一般概念：在幺半群范畴、cd-范畴或马尔可夫范畴中，具有给定查询集和语义的 *compositional model* ；因果模型及其查询（例如干预）是其中的特例。我们确定了两种基本的抽象概念，各自由自然变换描述：*downward abstractions*，适用于抽象图示查询，将它们从高层级映射到低层级；以及 *upward abstractions*，适用于具体查询（例如特定的Do-干预），将它们从低层级映射到高层级。尽管通常将常见的因果抽象概念表述为后一种形式，我们展示了如何从更根本的角度，依据前一种形式来理解它们。无论采用哪种形式，我们的范畴论方法都有助于凸显（因果）抽象的结构本质。此外，我们还提出了一种新的、更强的“组件级”抽象概念，它不仅适用于查询，也适用于模型的单个组件。具体而言，这带来了一种新颖且更强的 *constructive causal abstraction at the mechanism-level*形式，并且我们证明了其刻画结果。最后，通过将讨论推广到因果情形之外，我们展示了抽象也可以为其他组合模型定义，包括由量子电路实现 *quantum semantics* 的模型；我们并初步探索量子组合电路模型与高层经典因果模型之间的抽象，以此作为发展可解释量子人工智能的一种途径。

1 引言

为了依据具有解释力的高层级概念进行推理而忽略细节，是人类认知和科学实践的标志；从对气体的微观描述与热力学描述，到从气候数据或医学扫描中提取宏观变量，再到经济学，皆是如此。在所有这些情形中，以及几乎所有科学实践中，抽象问题都与因果论和因果推理问题交织在一起。

过去十年间，多项研究提出因果抽象这一概念，使这种直觉得以精确化 [CEP17, RWB+17, BH19, BEH20, MGIB23, GIZ+23]。这些研究以因果模型框架 [Pea09], 为基础，该框架将特定抽象层级上的因果模型与推理形式化。若一个“高层级”因果模型 $\mathbb{H}$ 与一个“低层级”因果模型 $\mathbb{L}$之间存在因果抽象关系，那么从低层变量到高层变量便存在抽象映射 τ，且各层级的干预之间存在某种关系，使得在高层级进行干预时，相对于低层级干预而言，因果上一致。

因果抽象在因果推断、人工智能和哲学等领域日益受到关注，既有基础层面的原因，也有实践层面的原因。尽管因果模型的假设、目标和形式各不相同，但在每种情境下都需要同一种基本关系。至少有三种不同情境推动了对它的研究。首先，它是对因果可识别性问题 的一种推广：我们的因果假设层级和关注的量，与我们获得经验数据的层级不同。在这种情况下，低层因果模型是一个真实模型（或其近似），而我们既无法获取它，通常也并不关心它。相关问题的例子涵盖从天气现象到人脑等不同领域

[CEP16、CEP17、XB24、HLT25]。此外，抽象对于种族歧视的因果研究也至关重要，有助于避免因错误混合不同抽象层级的关系项而导致的“模块性丧失” [MSEI25]。

其次，是对人工智能和深度学习中的**可解释性和解释性** 的研究。同样，其目标是将复杂且晦涩的低层模型（例如神经网络）与人类可理解的高层模型联系起来。不过，这里的低层模型是完全已知的，并且不被视为（对）真实模型（的近似），而是为某项任务训练得到的计算模型，该任务可能明确涉及因果关系，也可能不涉及；然而，与一个显然可解释的高层因果模型之间的抽象关系，可以提供一种作为黄金标准的可解释性。许多研究已将这一思路用于事后可解释人工智能方法 [GRP20, GLIP21, HT22, Bec22, WGI+24, GWP+24, GIZ+23, WGH+24, PMG25], with[GIZ+23] ，事实上，GIZ认为，**机制可解释性** 中的任何方法都最好从因果抽象的视角来理解。抽象也被用于非事后可解释人工智能，将其作为一种归纳偏置，以获得在设计上便于解释的模型 [GWL+22]。此外，有研究认为，抽象还可以更普遍地解释计算解释 [GHI25]。

第三个问题是**因果表征学习**，其目标是让人工智能模型直接从“原始”的低层级数据中学习适当的因果变量和关系，而不必（完全）预先知道这些变量的适当维度或数量。这一问题涵盖了因果发现以及其他形式的（解缠结）表征学习；有人认为，它对于人工智能的鲁棒性、强泛化能力和高效泛化能力至关重要 [SLB+21]。越来越多的研究探讨了这一问题在理论上和实践上是否（不）可能，且往往将因果抽象作为关键要素 [BDHLC22, YXL+23, KSB23, YRC+24, KGSL, LKR25, ZMSB24]。相关研究还表明，具身人工智能依赖于在恰当抽象层级上学得因果结构的真实世界模型 [GGM+24, RE24, REA25]。

一种用于抽象的语言。 尽管抽象是科学中的基础且不可或缺的概念，目前却尚无统一的因果抽象形式体系，也没有共识能够确定一系列密切相关的抽象概念中，哪些适用于何种情境。为了推进未来关于抽象的理论研究与实验研究，我们需要一种恰当的概念语言。

因果推理的核心在于句法因果结构与语义概率数据之间的相互作用，而抽象关注的是不同细节层级上的模型之间的关系。**范畴论**是研究过程、结构与复合的数学 [Coe06, Lei14]；它恰好是一种能将结构层面及其关系置于前景、从而直接加以研究的语言。事实上，范畴天然适合描述模型的结构，例如通过**幺半范畴** 和直观的**弦图** 图形语言 [PZ25, LT23, TLC+24]，也适合描述模型之间的关系，例如通过经典的**函子** 和**自然变换** 概念 [ML13]。

使用弦图以范畴论术语表述因果模型框架的方法，见于[LB21], ，其建立在 [JKZ19, FK23]的基础之上。在 [TLC+24] 中，这种方法扩展到了更一般的**组合模型**，并通过从结构到语义的函子来刻画，展示了多种人工智能模型（包括因果模型）的结构与可解释性均可用这些术语表达。

在本文中，我们扩展这一方法，对因果抽象进行精确且一般的形式化，并发现，抽象的本质可以更一般地在给定查询集合的组合模型层级上表达。更准确地说，我们表明，可以用模型之间的自然变换，从范畴论的角度对抽象进行形式化：

> 抽象是查询之间的自然变换。

具体而言，我们的贡献如下：

- **抽象的抽象定义**：我们提出了一个一般定义，用于描述给定查询集合的组合模型之间的抽象，并表明该定义涵盖了文献中许多因果抽象概念作为特例，例如构造性抽象 [BH19], 精确变换 [RWB+17]、“$Q-\tau$- 一致性” [XB24], 以及“分布式”抽象 [GIZ+23、GWP+24]。

- **图表的力量**：我们展示了如何将每种（因果）抽象的一致性条件直观地表示为弦图，从而凸显其本质。

• 上下之间：我们发现，实际上存在两种基本的抽象：一种将查询从低层映射到高层，另一种则从高层映射到低层；我们还发现，尽管许多典型的因果抽象概念通常以第一种“具体”类型呈现，其本质却在于结构性更强的后一种类型。

• 一般干预：我们给出因果抽象的统一处理，允许更一般的干预，而不必局限于Do-干预，并使我们能够给出分布式因果抽象的一种新的精确概念（与例如 [MGIB23、GIZ+23]保持一致）。

• 一种新的强概念：我们引入组件级 抽象，它们从范畴视角来看是自然的，并且在因果情形下产生一种新的更强的机制层面 构造性抽象概念，这提出了新的问题，我们对其作出数学刻画。[1]

• 量子推广：将抽象扩展到组合模型，使我们能够为量子组合模型（例如基于量子电路的模型）引入抽象的定义，并以此开启面向可解释量子人工智能的量子抽象研究。

相关工作。 若干工作已形式化定义了不采用范畴论视角的特定因果抽象概念，我们的观点主要基于 [RWB+17、BH19、GIZ+23、XB24]，其中我们重点关注后三者。（另见 [Zen22、Zha24]的综述。）

我们的方法建立在将结构化（人工智能）模型视为由弦图描述的组合模型这一论述之上；例如，（量子）自然语言处理中的 DisCo 模型 [CSC10、Coe21、WMLC23] 以及“discopy”和“Lambeq”Python 工具包 [dFTC20、KFY+21]都以此为基础。该框架中的因果模型论述建立在 [JKZ19、FK23], 之上，并已在 [LT23]中详尽阐述。 [TLC+24] 提出了将人工智能模型广泛视为组合模型的观点，以及组合结构对于可解释人工智能的优势（另见[dF22]）。在此，我们进一步拓展这一方法：除了将模型描述为函子之外，我们现在还将模型之间的关系描述为自然变换，并由此看到，这确实能够导出文献中独立出现的、有意义的抽象概念。

将因果抽象视为马尔可夫函子意义下的因果模型之间的自然变换这一思想，也独立地体现在 Englberger and Dhami [ED25], 的工作中，我们推荐读者进一步阅读。经过必要的修正后，他们的定义实际上对应于机制层面构造性因果抽象这一特殊情形，具体内容将在附录 C.1 中详细说明。我们基于查询的方法能够揭示构造性抽象与其机制层面版本之间的差异；结合我们关于向上抽象的概念，还可以统一更广泛的其他抽象形式，包括反事实抽象、精确变换、完全形式化的分布式抽象以及量子推广。

在 [OS22] 中，作者还利用自然变换研究了一种更为受限的因果抽象形式，但它最适用于“等价”模型，而非真正的抽象。我们的方法严格更为一般，并能刻画已知的因果抽象形式。

同样相关的是 [RW21], 中的工作，该工作以富集范畴的语言研究抽象关系的复合，尤其是误差界在复合下如何表现。

最后，一系列工作 [DLB25, DB25, DZFDL25, DDLB26], 除其他内容外，也用自然变换对某一类结构因果模型之间的抽象进行了形式化。然而，他们将因果模型形式化为函子的做法不同于我们的做法，即不同于 [LT23, JKZ19, FK23], 的做法；他们所刻画的抽象概念亦是如此，即 α-抽象 [Ris20]。在未来工作中，更充分地理解与这一方法的联系将是有意义的。

1.1 概述

现在让我们从高层级概述一下我们对抽象的论述。回顾一下，对称幺半范畴 $\mathbf{C}$ 由一组对象及其间的过程构成，我们可以用图形化弦图将这些过程复合起来。对于因果模型，我们使用范畴 $\mathbf{FStoch}$ ，它由有限集及其间的概率通道（随机矩阵）构成，并形成一个带有特定“复制”和

[1]后者隐含地出现在 [ED25]中；参见相关工作和附录 C.1。

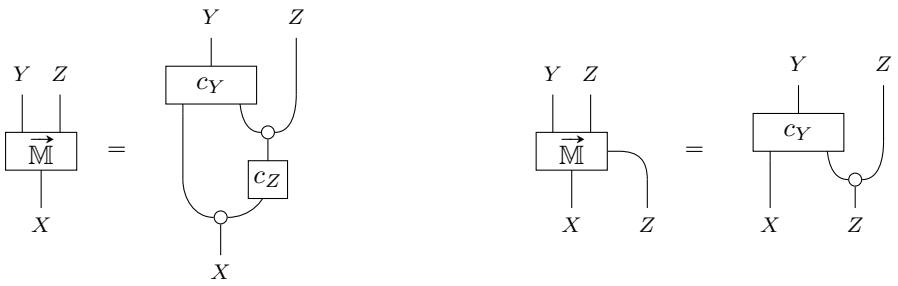
“丢弃”过程 [Fri20] 的马尔可夫范畴。在第2节中，我们将介绍范畴论和弦图方面的必要背景知识。

模型与查询。 组合模型 $\mathbb{M}$ 将语义赋予$\mathbf{C}$ 中的一组变量 和组件，其中每个组件都是一个抽象过程，并附有指定的输入变量和输出变量列表。每个变量 V 都表示为一个对象 $V = [\mathrm{V}]_{\mathbb{M}}$，每个组件 c 则表示为$\mathbf{C}$ 中的一个过程 $c = [c]_{\mathbb{M}}$（请注意，从语法转向语义时字体发生了变化）。对于因果模型， V是变量 V 的有限值集，每个 $c = c_V$则是因果机制 $P(V \mid \mathrm{Pa}(V))$的概率通道。形式上，变量和组件构成一个签名 $\mathcal{S}$，它生成一个自由范畴$\mathbf{S}$；模型则是一个函子$[-]_{\mathbb{M}}\colon \mathbf{S} \to \mathbf{C}$。

除此之外，模型还为一组查询赋予语义，用以描述我们在实践中如何使用模型。每个查询都会映射到$\mathbf{C}$中的一个过程 $Q = [\![\mathrm{Q}]\!]_{\mathbb{M}}$，通常由组件图表给出。形式上，这些查询构成另一个签名 $\mathcal{Q}$，其对象称为类型，而模型定义了一个函子 $[\![-]\!]\colon \mathbf{Q} \to \mathbf{C}$。

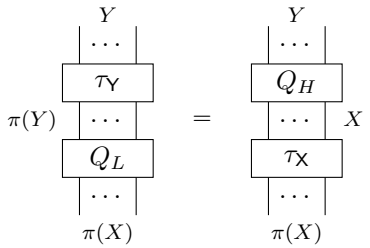
对于一个以变量 V $\mathbb{M}$ 为输入、以变量 V^{in} 为输出的因果模型，有各种标准的查询示例。一个基本示例是查询 $\vec{\mathbb{M}} = P(\mathrm{V}^{\mathsf{out}} \mid \mathrm{V}^{\mathsf{in}})$，以获得从输入到输出的整体概率通道。另一个是查询 $P(\mathrm{V}^{\mathsf{out}} \mid \mathrm{V}^{\mathsf{in}}, \mathrm{Do}(Z))$，它针对变量 Z 上一个尚待指定的 Do-干预，并将 Z 作为额外输入；为此，我们引入下方右侧所示的图形约定。

对于一个以 X 为输入、以 Y、Z 为输出，并具有因果关系 X →Y、Z 以及 Z →Y 的模型，查询 $P(Y, Z \mid X)$ 和 $P(Y, Z \mid X, \mathrm{Do}(Z))$ 分别由以下图表表示，阅读顺序为自下而上。



在第3节中，我们定义组合模型和查询，并将因果模型阐释为一种特殊情形。

抽象。 给定一个“低层级”模型 $\mathbb{L}$ 和一个“高层级”模型 $\mathbb{H}$，抽象关系首先为每个高层类型 X（或类型列表）指定一个低层类型列表 $\pi(\mathrm{X})$，以及一个“满射”（满态射）$\tau_{\mathsf{X}}\colon \pi(X) \to X$；直观而言，它将低层值映射到高层值。其次，它将每个高层级查询 Q_H 与一组低层查询 Q_L 联系起来，使得下述条件在$\mathbf{C}$中成立。



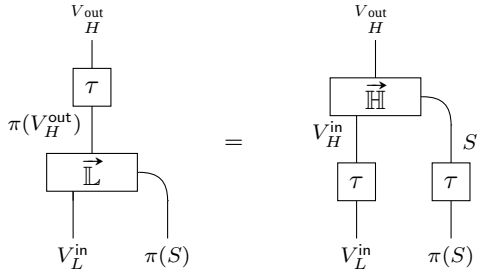
(1)

上述一致性 方程是抽象的基本规则。它规定，在抽象映射的意义下，各模型的行为是一致的：我们可以在低层级输入上实现每一个高层查询，既可以在 τ之后直接实现，也可以先应用相应的低层查询，然后再 τ。一个关键的观察是，从形式上讲，(1) 恰好表明映射 τ 构成了表示这些查询的函子之间的自然变换。抽象主要有两种类型，取决于关系 (Q_H、Q_L)的类型。

1. 一种*dow* ***nward*抽象** 有一个映射 $\pi\colon \mathcal{Q}_H \to \mathcal{Q}_L$，从高层查询到低层查询$\mathrm{Q}_L = \pi(\mathrm{Q}_H \quad)$.
2. 一个向上抽象具有一个偏映射 $\omega\colon \mathcal{Q}_L \to \mathcal{Q}_H$，将低层级查询映射到高层级查询 $\mathrm{Q}_H = \omega(\mathrm{Q}_L)$。

在第4节中，我们给出了模型之间抽象的完整定义。

一个关键示例是构造性因果抽象，它是因果模型之间的向下抽象，涉及形如 $P(\mathbf{V}^{\text{out}} \mid \mathbf{V}^{\text{in}}、\text{Do}(S))$ 的查询，并建立低层级 $\text{Do}(\pi(S))$ 与高层级 $\text{Do}(S)$ 上 Do-干预之间的一致性。根据上文的约定，一致性条件可写为：

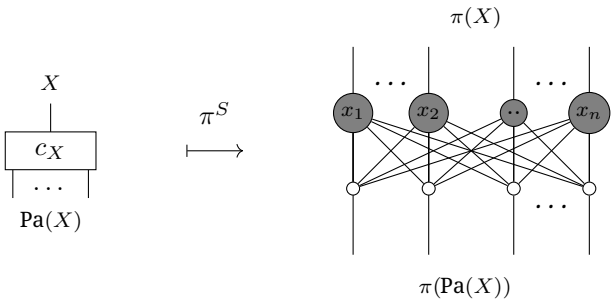


对于由特定 Do-干预 $P(\mathbf{V}^{\text{out}} \mid \mathbf{V}^{\text{in}}、\text{Do}(X = x))$ 给出的查询，通常有许多低层查询可以实现给定的高层查询。基于此类具体干预的抽象，无论是 Do- 风格还是更一般的形式，都由向上抽象来描述，在文献中称为“精确变换”。一般而言，对于结构层级上的任何“抽象”向下抽象，都有相应的“具体”向上抽象概念；我们在命题18中对此作了形式化表述。事实上，许多已知的因果抽象概念都可以用这两种形式来刻画。

第5节详细讨论因果抽象；它梳理了文献中的关键概念，指出这些概念是第4节定义的特例，并进一步发展了因果抽象理论。这些概念包括精确变换（见5.1）、构造性抽象（见5.2）、交换抽象（见5.3）、反事实抽象（见5.4），以及对 [GIZ+23] 中隐含讨论的“分布式”抽象的形式化说明（见5.5）。

组件级抽象。 范畴论视角启发我们考虑另一种更强的抽象形式，它不仅作用于查询，也直接作用于组件。所谓组件级抽象，是指配备了额外函子 π^S 的向下抽象：$\mathbf{S}_H \to \mathbf{S}_L$，它将每个高层级组件（例如因果机制）c 映射为一个低层级组件图表 $\pi^S(\text{c})$，并配有组件级函子之间的自然变换 τ，且该变换以恰当的方式保持查询。

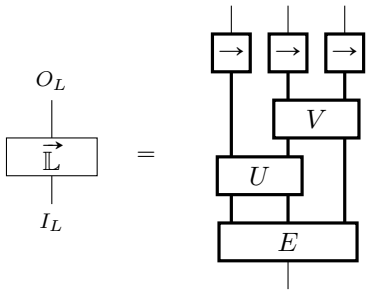
在严格 组件级抽象的最简单情形中，我们有 $\tau = \text{id}$，因此 $c = \pi^S(c)$ 属于 $\mathbf{C}$，只需将每个高层级组件写成一个等价的低层级组件图表即可。因果领域的例子包括将因果贝叶斯网络细化为结构因果模型（SCM），或直接用神经网络实现每个机制：



更一般地，对于一个（不一定严格的）组件级抽象，原始组件 c和图表 $\pi^S(c)$ 必须满足$\mathbf{C}$中的自然性条件 $c \circ \tau = \tau \circ \pi^S(c)$。我们将以这种方式扩展为组件级抽象的构造性因果抽象称为机制级因果抽象。我们认为这是一个新概念，并在定理 51 中证明，它对应于构造性抽象的划分 π 所满足的某些条件。在未来工作中，探究机制级因果抽象的意义，以及它们在典型抽象示例中的适用频率，将会很有意思。

第 6 节介绍了组合模型的组件级抽象这一新概念，并特别讨论了第 6.2 节中的机制级因果抽象。

量子抽象。 在组合模型层级定义抽象，使其能够应用于“经典”因果模型之外的情形，尤其是具有量子 语义的组合模型。现在，我们可以考虑这样一种情况：低层级组合模型 $\mathbb{L}$ 的输入输出过程由一个量子电路给出，例如下面这个简单示例，其中细线表示经典输入和输出，粗线表示量子系统，如量子比特。这里， E 是一个将经典输入编码为量子比特的编码器，U、V是幺正门，上方的盒子表示测量。



还可以将（抽象）Do-干预以及较弱的“互换”干预推广到由此类电路定义的量子模型。然后，我们可以基于任何此类查询，考察量子电路模型 $\mathbb{L}$ 与经典因果模型 $\mathbb{H}$ 之间的抽象关系。此类抽象或许能让我们像在经典可解释人工智能中运用抽象那样，对量子电路的某些方面给出经典的（因果）解释 或阐释；例如，在量子人工智能场景中，电路可能作为训练产生的“黑箱”出现。

在第 7 节中，我们研究量子组合模型的抽象问题，其动机是量子人工智能中的解释性。最后，我们在第 8 节讨论未来的研究方向。附录中收录了一些关于该理论及部分证明的补充材料。

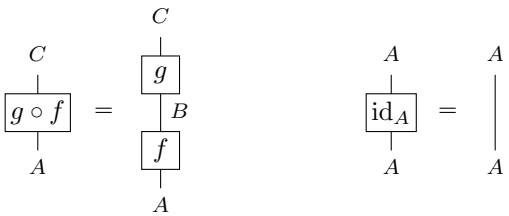
2 范畴与弦图

我们将运用范畴论及其相应的图形语言——弦图 [Coe06, PZ25, Sel11]。具体而言，全文将使用一个对称幺半范畴（对称幺半范畴） $(\mathbf{C}, \otimes)$ [Coe06]。

回顾一下，范畴由一组对象 A、B、$\ldots$ 以及它们之间的态射或过程 $f\colon A \to B$构成。在弦图中，对象 A 表示为一条线，态射表示为一个盒子，按从下到上的方向阅读。

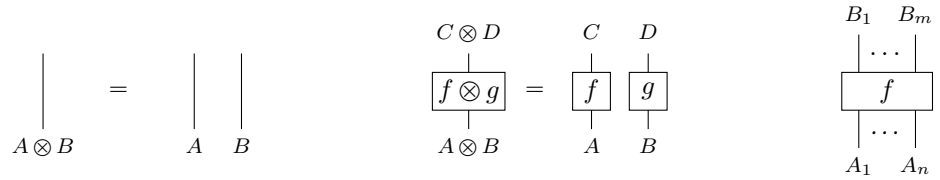


对于任意态射 $f\colon A \to B$和 $g\colon B \to C$，我们可以构造它们的顺序复合 $g \circ f\colon A \to C$，如下图所示。对象 A 上的恒等态射 id_A，满足对任意 $f\colon A \to B$ 都有 $\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A$，表示为一条空白线。

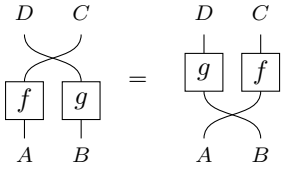


在幺半范畴中，对于任意一对对象 A、B，我们都可以构造它们的幺半积 $A \otimes B$，其恒等态射由下方左侧所示的并排导线表示。更一般地，对于态射 $f\colon A \to C$ 和 $g\colon B \to D$，我们可以构造它们的幺半积 $f \otimes g\colon A \otimes B \to C \otimes D$，如下方中间所示。一般而言，图表中的态射可以有多个输入和/或输出，这在形式上意味着

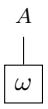
它们的输入以这样的乘积给出。例如，我们将态射 $f\colon A_1 \otimes \cdots \otimes A_n \to B_1 \otimes \cdots \otimes B_m$ 表示为下方右侧所示的形式。



由于我们的范畴是对称的，我们还有允许将成对导线彼此交叉交换的态射，使得 $\circ$ =id，并且盒子可以沿着交换滑动：



请记住，总是存在一个 单位对象 I，它简单地表示为“空白空间”；与 I 取幺半积只会使任意对象保持不变。[2] 我们写作 $1 := \mathrm{id}_I$。一类值得注意的态射是 A的状态，它们是态射 $\omega\colon I \to A$，以“无输入”的形式出现。



范畴中的复合还满足许多其他公理，这些公理在字符串图中显而易见；例如，复合的结合律 $(h \circ g) \circ f = h \circ g \circ f = h \circ (g \circ f)$ 由以下事实保证：三个复合项都相当于依次绘制 f、g、h。

定义 1 (终端范畴)。 *[Coe08, Coe14]* 当一个对称幺半范畴 $\mathbf{C}$ 满足 I 是终对象时，就称其为终端的。也就是说，对于每个对象 A，都存在唯一的态射 $A \to I$，我们称之为丢弃 $_A$，并用一个接地符号表示：

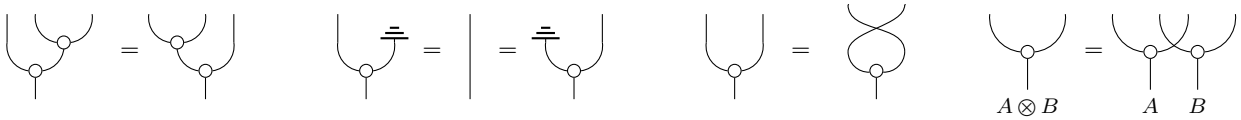


我们将通过丢弃态射 f a 的任意输出而得到的态射称为 f 的边缘分布。

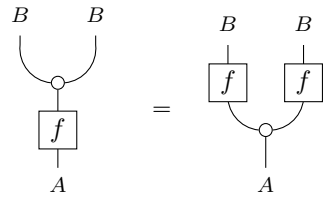
我们也将终端范畴中的态射称为通道。

本文中的主要范畴示例都将是终端范畴。我们将使用的最后一种额外结构存在于“经典”（相对于量子）过程中，例如因果模型中的过程，具体如下。

定义 2 (马尔可夫范畴)。 *[Fri20]* A 马尔可夫范畴是一个终端的对称幺半范畴$\mathbf{C}$，其中每个对象 A都配备了一个选定的复制态射： $A \to A \otimes A$，并满足：



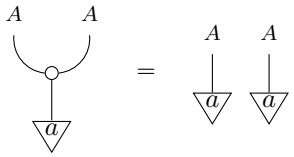
在这 样的范畴中，当态射 $f\colon A \to B$ 确定性的满足以下条 g.



(2)

[2]形式上，这是通过“相干同构” $A \otimes I \simeq A \simeq I \otimes A$来表达的。

I特别地，当状态 a 被以下映射复制时，它就是确定性的或尖锐的：



(3)

对于因果模型，本文只需要以下范畴。

例 3 (**FStoch**)。在马尔可夫范畴 **FStoch** 中，对象是有限集 $X, Y, \ldots\ldots$。态射$f\colon X \to Y$是从 X到 Y的概率通道，为每个 $x \in X$和$y \in Y$指定一个概率 $f(y \mid x) \in [0,1]$，使得：

$$\sum_{y \in Y} f(y \mid x) = 1$$

对所有 $x \in X$均成立。对于任意 $f\colon X \to Y$和 $g\colon Y \to Z$，顺序复合 $g \circ f\colon X \to Z$由矩阵复合给出：

$$(g \circ f)(z \mid x) := \sum_{y \in Y} g(z \mid y) f(y \mid x)$$

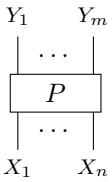
X 上的恒等通道由 $\mathrm{id}(y \mid x) = \delta_{x,y}$给出。幺半积在对象上的定义为$X \otimes Y = X \times Y$；对于通道 $f\colon X \to W$和 $g\colon Y \to Z$，其定义为

$$(f \otimes g)(w, z \mid x, y) = f(w \mid x) g(z \mid y)$$

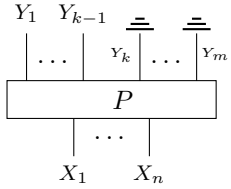
对应于两个独立概率通道。这里 $I = \{\star\}$是一个单元素集，因此， X上的状态 ω 可通过 $\omega(x) := \omega(x \mid \star)$对应于 X 上的分布。

X 上的复制态射是 $(y,\ z \mid x) = \delta_{x,\ y,\ z}$。当且仅当某个 $x \in X$ 对应的点分布为 δ_x 时，状态才是尖锐的。我们将这样的状态简记为 x，因此 X 的尖锐状态 x 与 X 的元素相对应。更一般地，确定性态射 $f\colon X \to Y$ 通过 $f \circ x = f(x)$与函数 $f\colon X \to Y$ 相对应。丢弃 X 上的 X，对应于唯一的函数 $X \to I$。

概率论中的许多概念都可以用 **FStoch** 和更一般的马尔可夫范畴中的弦图来描述 *[Fri20]*。[3] 特别地，如下的一个态射正是一个条件概率分布 $P(Y_1, \ldots, Y_m \mid X_1, \ldots, X_n)$：



而与丢弃的复合相当于通常意义上的边缘化：

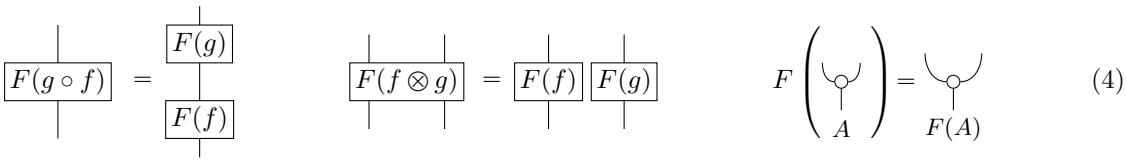


$$= \sum_{\substack{y_k, \ldots y_m \in \\ Y_k \times \ldots \times Y_m}} P(Y_1, \ldots, Y_{k-1}, Y_k = y_k, \ldots, Y_m = y_m \mid X_1, \ldots, X_n)$$

函子与自然变换。下文中，范畴始终取自某个固定的范畴类，即终对称幺半范畴或马尔可夫范畴。回顾一下，范畴$\mathbf{C}$与$\mathbf{D}$之间的一个函子 F由对象和态射上的映射给出 $A \mapsto F(A)$and $(f\colon A \to B) \mapsto F(f)\colon F(A) \to F(B)$从$\mathbf{C}$ 到$\mathbf{D}$，并保持复合与恒等态射。在这里，我们始终指强对称的函子

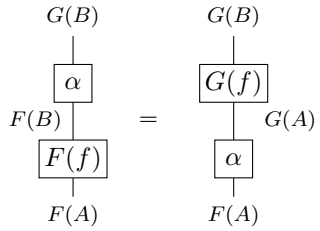
[3]进行这类概率推理时，在更广泛的范畴$\mathbf{Mat}_{\mathbb{R}^+}$ 中进行研究也可能很有用。该范畴由正实矩阵构成，而 **FStoch** 是其中通道的子范畴，例如参见 [LT23]；此处后一个范畴已足够。

幺半的，即保持 I、$\otimes$ and 。在马尔可夫情形下，我们还始终要求函子保持复制态射，并具有确定性结构同构。[4]



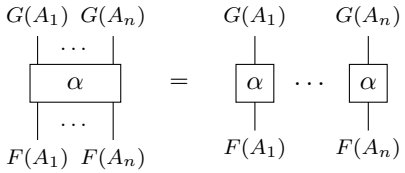
(4)

给定函子 $F, G\colon \mathbf{C} \to \mathbf{D}$，一个自然变换 $\alpha\colon F \Longrightarrow G$ 由一族态射 $\alpha_A\colon F(A) \to G(A)$ 给出。在图表中，由于线标签的存在，这些态射都简写为 α；对于所有 $f\colon A \to B$，都有：



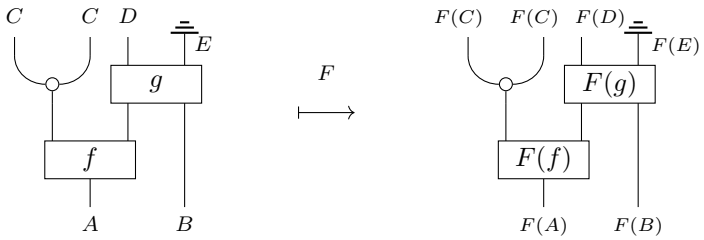
(5)

我　　们总是默认指幺半群的自然变换，即：[5]

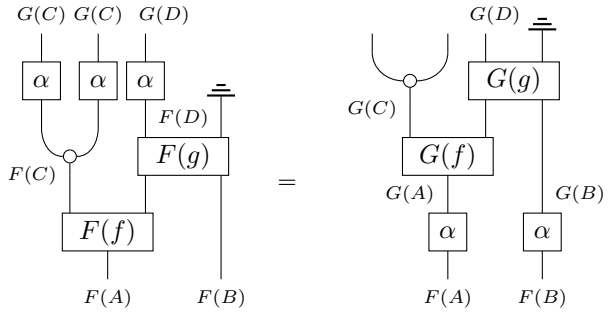


并且在马尔可夫情形下，每个 α_A 都是确定性的。

注 4. 直观地说，对于每种范畴，相应类型的函子 $F\colon \mathbf{C} \to \mathbf{D}$ 会将 $\mathbf{C}$ 中的任意图表映射到 $\mathbf{D}$ 中形状相同的图表，并保留.....（若有）。



自然变换 α 具有必要的性质，使我们能够将它通过如下形式的任意此类图表“拉过”：$F(f)$, $G(f)$。



[4]更具体地说，函子满足 $F(g \circ f) = F(g) \circ F(f)$ 和 $F(\mathrm{id}) = \mathrm{id}$。强幺半函子附带一个同构 $F(I) \simeq I$，以及满足相干条件的自然同构 $F(A \otimes B) \simeq F(A) \otimes F(B)$。马尔可夫函子满足这样的性质：方程 (4) 在 $\otimes$ 和.....的情况下，在结构同构意义下成立。在图表中，我们省略结构同构的绘制。[5]此外，I，因为 $\alpha_I\colon F(I) \to G(I)$ 是 F, G 的结构同构的复合。

3 组合模型与查询

现在，我们可以定义本文将要讨论的模型的一般形式。首先，我们从指定其抽象结构开始，如下所示。

定义 5 (签名)。 一个 (幺半)签名 $\mathcal{G}$ 由以下部分组成：

- 一个集合 $\mathcal{G}_{\mathsf{ob}}$，其中包含抽象的“对象”A、B、C、.....；
- 一个集合 $\mathcal{G}_{\mathsf{mor}}$，其中包含抽象的“态射”f: $(\mathsf{A_i})_{i=1}^n \to (\mathsf{B_j})_{j=1}^m$，其中每个态射都附有一个由输入对象 $(\mathsf{A_i})_{i=1}^n$ 和输出对象 $(\mathsf{B_j})_{j=1}^m$ 构成的列表，这些对象来自$\mathcal{G}_{\mathsf{ob}}$。

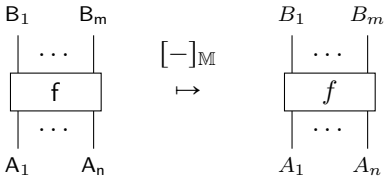
我们固定某一类幺半范畴及其关联的弦图，选取终对称幺半范畴或马尔可夫范畴。[6] 任意签名 $\mathcal{G}$ 都会自由生成一个此类范畴 $\mathbf{G}$ ： 对象是来自$\mathcal{G}_{\mathsf{ob}}$ 的列表，态射 $\mathbf{D}$: $(\mathsf{A_i})_{i=1}^n \to (\mathsf{B_j})_{j=1}^m$ 是由 $\mathcal{G}_{\mathsf{mor}}$ 构成的所选类型的弦图，其输入为$(\mathsf{A_i})_{i=1}^n$ ，输出为 $(\mathsf{B_j})_{j=1}^m$。 [7] 如果两个图表 $\mathbf{D}$、$\mathbf{D}'$ 能够通过有限步使用此类图表的规则相互改写，我们就认为它们作为态射相等。

我们现在可以定义在一个范畴中对这种结构进行建模意味着什么。以下概念在本文中居于核心地位。

定义 6 （模型）。 签名 $\mathcal{G}$ 在语义 范畴 $\mathbf{C}$ 中的模型 $\mathbb{M}$ 由一个函子构成：

$$[-]_{\mathbb{M}} \colon \mathbf{G} \to \mathbf{C}$$

等价地，这相当于为每个 $\mathsf{A} \in \mathcal{G}_{\mathsf{ob}}$ 指定 $\mathbf{C}$ 中的一个对象 $A := [\mathsf{A}]_{\mathbb{M}}$，并为 $\mathcal{G}_{\mathsf{mor}}$ 中的每个 f 指定 $\mathbf{C}$ 中一个具有如下适当类型的态射 $f := [\mathsf{f}]_{\mathbb{M}}$。



(6)

换言之，一个模型在$\mathbf{C}$中选取与 $\mathcal{G}$ 中“抽象”对象和态射相对应的具体对象和态射。如上所述，我们通常省略不写 $[-]_{\mathbb{M}}$，并使用不同字体来区分 $\mathbf{G}$ 中的任意对象 $\mathbf{A}$ 或态射 $\mathbf{f}$ 与其在 $\mathbf{C}$ 中的语义 A、f。

我们在此考虑的模型实际上通常带有两个 不同但相关的签名。尽管二者都只是上述定义的特例，为它们引入不同的术语仍有帮助。

定义 7 （组合模型）。 所谓组合模型 $\mathbb{M}$，是指一个签名 $\mathcal{S}$，连同 $\mathcal{S}$ 在语义 范畴 $\mathbf{C}$ 中的一个模型：

$$[-]_{\mathbb{M}} \colon \mathbf{S} \to \mathbf{C}$$

其中，我们将模型的 $\mathcal{S}$ 结构称为， $\mathcal{S}_{\mathsf{ob}}$ 称为变量，而模型的 $\mathcal{S}_{\mathsf{mor}}$ 组件称为。

虽然这只是术语上的说法，但将一个模型称为组合模型，是为了表明其签名$\mathcal{S}$ 体现了组合结构，即所建模现象中关注的基本底层组件、机制或构件过程。通常还会关注由这些构造出的其他过程，而模型的目标就是计算这些过程；例如，模型需要回答的特定问题。我们将这些称为“查询”，并按如下方式将其组织到另一个签名中。

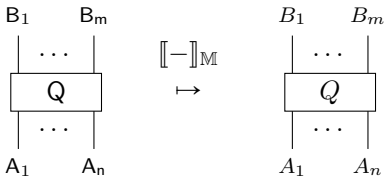
[6]在第 6.2 节中，我们还会讨论 cd 范畴和笛卡尔范畴。[7]复合就是在“页面上”复合图表，空列表是单位元，而每个恒等映射都是由普通线构成的图表。签名还可以包含一组 $\mathcal{G}_{\mathsf{eq}}$ $equations$ $\mathsf{D}_1 = \mathsf{D}_2$，其中 D_1,D_2 是由生成元构成的弦图，但我们在这里不重点讨论这一点。当图表可以通过使用 $\mathcal{G}_{\mathsf{eq}}$相互重写时，它们是等价的。模型应保留 $\mathcal{G}_{\mathsf{eq}}$中的任何方程。

定义 8（查询模型）。我们将其态射称为查询签名，简单来说，就是一个签名 $\mathcal{Q}$ 。我们将 $\mathcal{Q}$ 中的对象（列表）称为查询类型。我们将 $\mathcal{Q}$ 在$\mathbf{C}$ 中的一个模型 $\mathbb{M}$ 简称为查询模型，其函子记作：

$$[\![-]\!]_{\mathbb{M}} \colon \mathbf{Q} \to \mathbf{C}$$

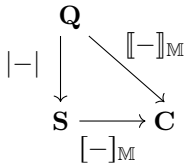
所谓组合式查询模型 $(\mathbb{M},\ \mathcal{S},\ \mathcal{Q},\ \mathbf{C})$，简单来说，就是一个结构的组合模型 $[-]_{\mathbb{M}}$在$\mathcal{S}$ $\mathbf{C}$中，以及一个查询模型 $[\![-]\!]_{\mathbb{M}}$ 在$\mathbf{C}$中。

因此，同样地，每个查询都会指定一个抽象输入输出过程 Q，而模型 $\mathbb{M}$ 赋予其语义$Q = [\![\mathsf{Q}]\!]_{\mathbb{M}}$，并通过改变字体来表示。



通常，我们会使用一个组合式查询模型 $(\mathbb{M},\mathcal{S},\mathcal{Q},\mathbf{C})$，而不只是一个查询模型。当然，同一组查询 $\mathcal{Q}$可能有许多组合模型，但它们的结构 $\mathcal{S}$各不相同。此外，对于任意这样的元组 $(\mathbb{M},\mathcal{S},\mathcal{Q},\mathbf{C})$，这两个函子通常并非任意的，而是紧密相关；它们之间一种尤为紧密的联系是可以按如下方式分解。

定义 9（抽象查询）。对于任意组合式查询模型 $(\mathbb{M},\ \mathcal{S},\ \mathcal{Q},\ \mathbf{C})$，如果这些查询$\mathcal{Q}$ 是抽象的，即它们配有一个函子 $|-| : \mathbf{Q} \to \mathbf{S}$，将每个查询映射到 $\mathbf{S}$中的一个图表，并使下图交换，则称这些查询为抽象查询。

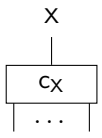


给定一个查询的组合模型 $(\mathbb{M}、\ \mathcal{S}、\ \mathcal{Q}\mathbf{C})$，即使查询在技术上并非抽象查询，每个查询 $[\![Q]\!]_{\mathbb{M}}$ 的计算通常也会使用该组合模型的结构 $\mathbf{S}$和语义 $[-]_{\mathbb{M}}$ 。最后，当上下文中已明确具体模型 $\mathbb{M}$ 时，我们可以省略下标，直接分别用 $[-]$ 和 $[\![-]\!]$ 表示结构与查询的语义函子。

3.1 因果模型

现在，让我们通过组合模型和查询的主要示例——因果模型——来说明我们的设置。下面的介绍将紧密遵循 [LT23]（以及 [JKZ19]）的论述。

所谓开放有向无环图 $G = (G,\ \mathsf{V}^{\mathsf{in}})$，是指一个定义在顶点 V 上的有向无环图 G，以及顶点 V 的子集 $\mathsf{V}^{\mathsf{in}} \subseteq \mathsf{V}$，该子集中的每个输入顶点在 G中都没有父节点。随后，我们定义一个签名 $\mathcal{S}_G$：对每个顶点 $\mathsf{X} \in \mathsf{V}$，都包含一个对象，即变量 $\mathbf{X}$；并且对每个非输入顶点 $X \in \mathsf{V}^{\mathsf{in}\perp}$，包含一个生成元$= \mathsf{V}\backslash\mathsf{V}^{\mathsf{in}}$，称为 $\mathbf{X}$ 的（因果）机制：



$\mathrm{Pa}(\mathsf{X})$

其中，$\mathrm{Pa}(\mathsf{X})$ 表示 G 中 $\mathbf{X}$ 的父节点。我们将 $\mathbf{S}_G$ 记作由 $\mathcal{S}_G$ 生成的自由马尔可夫范畴，即由这些生成元以及复制映射和丢弃映射所能构造的所有图表组成的范畴。

定义 10. 设 $\mathbf{C}$ 是一个马尔可夫范畴，G 是一个开放有向无环图。$\mathbb{M}$ 中 G 在 $\mathbf{C}$ 中的一个（开放）因果模型由 $\mathbf{C}$ 中 $\mathcal{S}_G$ 的一个组合模型，以及对输出变量 $\mathsf{V}^{\mathsf{out}} \subseteq \mathsf{V}$[8] 的指定构成。

[8]输出变量也可以选择视为“观测到”的变量。

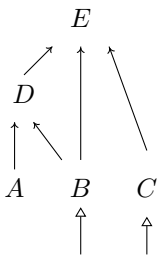
在 **FStoch** 中，一个因果模型 $\mathbb{M}$ 由 G 构成，即为每个变量 $\mathbf{X}$ 指定一个有限集 $X := [\![\mathbf{X}]\!]$ 的取值，并为每个非输入变量 $X \in \mathbf{V}^{\mathsf{in}\perp}$ 指定一个概率通道 $c_X \colon \mathrm{Pa}(X) \to X$，更常见的表示方式为：[9]

$$P(X \mid \mathrm{Pa}(X))$$

因此，这样的模型恰好是针对有向无环图 G 的一个（开放）因果贝叶斯网络（*CBN*），也是有限集上典型的因果模型概念。这里的"开放"指的是，G 可以带有输入，而这些输入没有指定机制。[10]

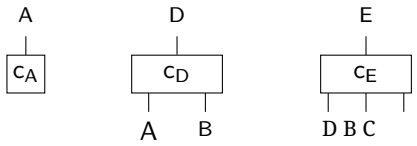
以下概念也会派上用场。所谓一组输入输出变量$\mathbf{V}$，是指一组（抽象）变量，并给定其子集 $\mathbf{V}^{\mathsf{in}}$、$\mathbf{V}^{\mathsf{out}} \subseteq \mathbf{V}$，分别作为输入变量和输出变量。所谓一组具体变量$\mathbb{V} = (\mathbf{V}, [\![-]\!])$ 属于$\mathbf{C}$，是指一组输入输出变量 $\mathbf{V}$，并为每个 $\mathbf{X} \in \mathbf{V}$ 给定一个$\mathbf{C}$ 中的语义 $X = [\![\mathbf{X}]\!]$。[11] 有时我们将 V 简写为 $\mathbb{V}$。于是，定义在 $\mathbb{V}$ 上的一个因果模型 $\mathbb{M}$，是指某个 $\mathcal{S}_G$ 上的因果模型，其变量为 $\mathbf{V}$，输入为 $\mathbf{V}^{\mathsf{in}}$，输出为 $\mathbf{V}^{\mathsf{out}}$，并且对所有 $X \in \mathbf{V}$ 都有 $X = [\![\mathbf{X}]\!]_{\mathbb{M}}$。

示例 11. 考虑开放有向无环图 G，其定义在$\mathbf{V} = \{\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}, \mathsf{E}\}$上，输入为$\mathbf{V}^{\mathsf{in}} = \{\mathsf{B}, \mathsf{C}\}$，如下方左端所示；其中我们用"无源"箭头来表示子集$\mathbf{V}^{\mathsf{in}}$。按常规表述，具有该因果结构的开放因果贝叶斯网络指定（有限）集合 A，B，C，D，E，以及右端列出的概率通道，还有被视为输出的变量子集，例如$\mathbf{V}^{\mathsf{out}} = \{\mathsf{C}, \mathsf{D}, \mathsf{E}\}$。

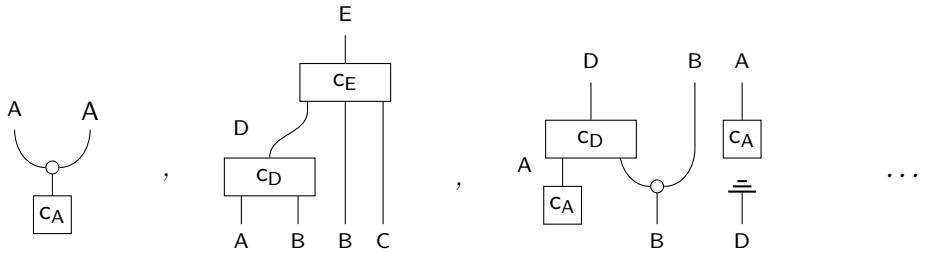


$$P(E|B,C,D)$$
$$P(D|A,B)$$
$$P(A)$$

用组合模型的语言来说，相应的签名 $\mathcal{S}_G$ 由对象（变量）A，B，C，D，E 以及态射（机制）组成：



马尔可夫范畴$\mathbf{C}$中 G 的因果模型 $\mathbb{M}$由函子 $[\![-]\!]$给出，该函子将每个变量映为一个对象$A = [\![\mathsf{A}]\!]$，将 $B = [\![\mathsf{B}]\!]$，……，并将每个机制映为其对应的通道；例如，对于$\mathbf{C} = \mathbf{FStoch}$，有$[\![\mathsf{c_D}]\!] = c_D = P(D \mid A, B)$。由 $\mathcal{S}_G$ 生成的自由马尔可夫范畴$\mathbf{S}_G$ 以态射的形式包含我们可由这些组件构建的所有马尔可夫图，例如：



The 函子 $[\![-]\!]$ 会将任何此类图表 D 映射为一个通道 $[\![\mathsf{D}]\!]$，从其输入映射到其输出 *s.*

[9]对于$\mathbf{C} = \mathbf{FStoch}$，我们采用 $P(-|-)$的常用记法，但机制 c_X 实际上就是一个随机通道，不必由联合分布的条件分布产生；另见 [LT23]。[10]尽管文献中并不常见从 [LT23] 引入的、具有输入的因果模型定义，但这种定义实际上非常自然，因为例如，"未指定"的外生变量或 Do-干预都会产生这类开放模型（见下文）。[11]这种记法反映出，$\mathbb{V}$ 在形式上是一个将 $\mathbf{V}$ 视为无态射签名的模型。具体变量 $\mathbb{V}$ 对应于因果模型文献中有时所称的"签名"。

3.2 因果查询

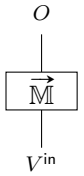
对于因果模型，可以有多种不同类型的查询，分别对应于我们可以提出的各种“因果问题”，通常基于干预。这些查询通过计算由模型机制导出的相应通道，在$\mathbf{C}$中得到“回答”（即建模）。现在我们将介绍因果模型查询的几个示例，从最基本的开始。

3.2.1 输入输出查询

最简单的做法是，将从输入 $\mathbf{V}^{\text{in}}$ 到输出的任意子集 $\mathbf{O} \subseteq \mathbf{V}^{\text{out}}$ 的整体通道视为一个查询。对于因果贝叶斯网络，这就是整体通道的边缘分布 $P(O \mid V_{\text{中}})$ 于 O 上的分布：

$$P(V \mid V^{\text{in}}) \;:=\; \Big(\prod_{X \in \mathbf{V}^{\text{in}\perp}} P(X \mid \mathrm{Pa}(X))\Big)\Big(\prod_{Y \in \mathbf{V}^{\text{in}}} P(Y \mid Y)\Big)\,, \tag{7}$$

第二个因子可以省略，因为 $P(Y \mid Y) = \delta_{Y,\,Y}$但保留它是为了强调：一般来说，我们允许 $\mathbf{V}^{\text{out}} \cap \mathbf{V}^{\text{in}} \neq \varnothing$，在这种情况下，相应的输入会被“复制出去”，并同时视为输出。更一般地，对于定义在 G 上的任意因果模型 $\mathbb{M}$ 以及输出子集 $O \subseteq \mathbf{V}^{\text{out}}$，$\mathcal{S}_G$ 包含一个称为（规范化）网络图 $\mathbf{D}_{\mathbb{M}}$的图表：$\mathbf{V}^{\text{in}} \to \mathbf{O}$。该图表通过复制映射，将每个具有通向 $\mathbf{O}$ 的路径的变量 $\mathbf{X}$ 的机制连接起来。每个 $\mathbf{X} \in \mathbf{V}^{\text{in}}$ 都是该图表的输入，而每个 $\mathbf{X} \in O$ 都会额外复制一次，形成该图表的输出。[12] 我们将所得通道记作 $\vec{\mathbb{M}} := [\mathbf{D}_{\mathbb{M}}]_{\mathbb{M}}$，它属于 $\mathbf{C}$：

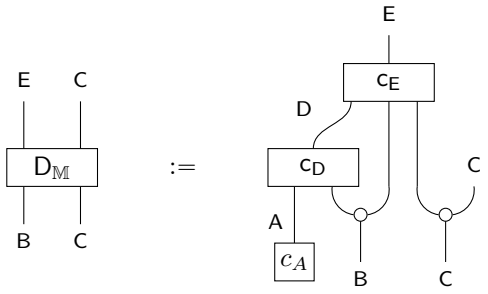


(8)

对于因果贝叶斯网络，$\vec{\mathbb{M}}$ 恰好就是概率通道 $P(O \mid \mathbf{V}^{\text{in}})$。示例如下，完整细节请参见[JKZ19、LT23、FL23]。

形式上，我们定义一个查询签名 $\mathrm{io}(\mathbf{V})$，其类型为 $\mathbf{V}$；并且对每个子集 $\mathbf{O} \subseteq \mathbf{V}^{\text{out}}$，定义一个查询 $(\mathbf{O} \mid \mathbf{V}^{\text{in}})$: $\mathbf{V}^{\text{in}} \to \mathbf{O}$。于是，$\mathbb{M}$ 通过 $|V| = V, \big|(O \mid \mathbf{V}^{\text{in}})\big| = \mathbf{D}_{\mathbb{M}}$ 将它们表示为抽象查询，使得 $[\![(O \mid \mathbf{V}^{\text{in}})]\!] = (8)$。

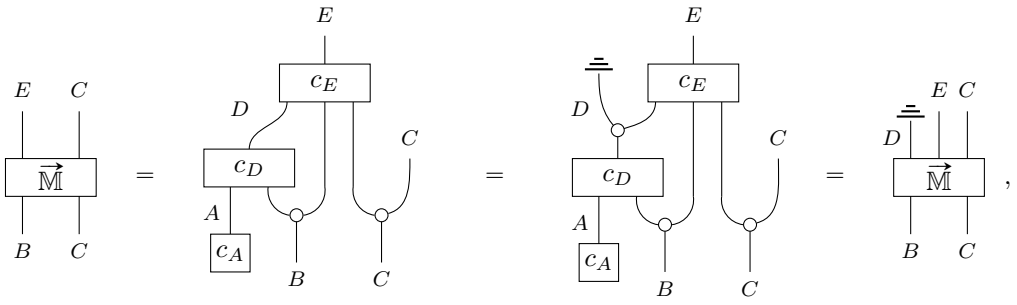
示例 12. 对于示例 *11* 中的有向无环图 G 和模型 $\mathbb{M}$，其针对$\mathbf{O} = \{\mathsf{E}, \mathsf{C}\} \subset \mathbf{V}^{\text{out}}$ 的网络图如下：



因此，$io(\mathbf{V})$ 中相应的查询 $(\mathbf{O} \mid \mathbf{V}^{\text{in}})$: $\mathbf{V}^{\text{in}} \to \mathbf{O}$ 被映射为等价的复合 $\vec{\mathbb{M}}$，其为

[12] 与 [LT23] 不同，这里我们使用规范化网络图，其中每个非输入变量都有一条通向输出的路径。特别地，这意味着在 G 中，任何没有通向 O 的输入 $X \in \mathbf{V}^{\text{in}}$，在 $\mathbf{D}_{\mathbb{M}}$ 中都是不连通的丢弃 $_X$ 的输入。请注意，为避免图表过于杂乱，在书写 $\mathbf{D}_{\mathbb{M}}$ 时省略 O 并无妨，因为在图表中，O 始终会作为类型明确标出。

c通道$\mathbf{C}$ 中的通道（注意字体变化表示转到$\mathbf{C}$中的特定通道）：

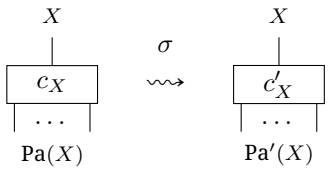


其中我们还展示了它如何作为 $\vec{\mathbb{M}}$ 中满的 $(\mathsf{V}^{\mathsf{out}} \mid \mathsf{V}^{\mathsf{in}})$ 的边缘分布而产生。对于 $\mathbf{C} = \mathbf{FStoch}$ 中的开放因果贝叶斯网络，我们有 $\vec{\mathbb{M}} = P(E, C \mid B, C)$，上述内容等价于以下公式（与式 (7) 相比，去掉因子 $P(C \mid C)$）：

$$P(E, C \mid B, C) = \sum_{a \in A, d \in D} P(E \mid D = d, B, C)\ P(D = d \mid A = a, B)\ P(A = a)$$

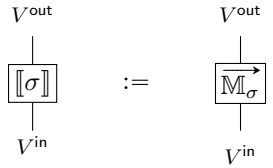
3.2.2 将一般干预作为查询

给定 $\mathbf{C}$ 中任意因果模型 $\mathbb{M}$，一个作用于 $\mathbb{M}$ 的干预 σ，是对一个新因果模型 $\mathbb{M}_\sigma$ 的规定；该模型具有相同的变量、输入和输出，包括其在 $\mathbf{C}$ 中的语义 $V = [\![\mathbf{V}]\!]$。我们可以将干预视为一个映射，它将新的机制 $\{c'_X\}_{X \in A_\sigma}$ 指派给变量子集 $A_\sigma \subseteq \mathbf{V}^{\mathsf{in}^\perp}$。[13]



(9)

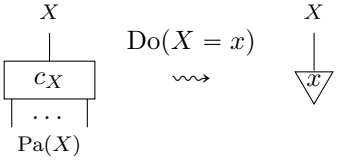
形式化地说，给定具体变量 $\mathbb{V}$，干预集 $\mathcal{I}$ 可视为一个查询签名，其类型为 $\mathbf{V}$，并且对每个 $\sigma \in \mathcal{I}$（对应于数据 $\{c'_X\}_{X \in A_\sigma}$）都有一个查询 $\sigma\colon \mathbf{V}^{\mathsf{in}} \to \mathbf{V}^{\mathsf{out}}$。任何在 $\mathbb{V}$ 上的因果模型 $\mathbb{M}$，只要所有 $\sigma \in \mathcal{I}$ 都是定义良好的干预（即不会引入有向环），就会通过以下方式产生 $\mathcal{I}$ 在$\mathbf{C}$ 中的一个模型：



为便于后文使用，我们指出，干预集 $\mathcal{I}$定义相应查询签名还有第二种仅略有不同的方式：[14] 令 $\mathcal{I}^{\mathsf{io}}$ 表示具有仅两种 类型 in、out 以及查询 σ的签名：对每个 $\sigma \in \mathcal{I}$，$\mathsf{in} \to \mathsf{out}$，其中 $[\![\mathsf{in}]\!] = V_{\mathsf{in}}$，$[\![\mathsf{out}]\!] = V_{\mathsf{out}}$。

3.2.3 Do 查询

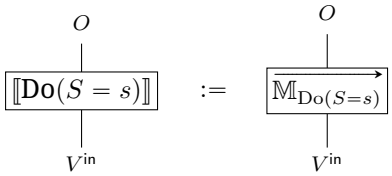
干预的典型形式是 Do-干预。给定具体变量 $\mathbb{V}$ 属于$\mathbf{C}$，*Do*-干预 $\mathrm{Do}(X = x)$ 作用于 $\mathbf{X} \in \mathbf{V}^{\mathsf{in}^\perp}$ 是如下形式干预的特例：



[13]鉴于本文聚焦于无环因果模型，规定对给定模型进行干预的新机制 $\mathbb{M}$ 不受其他限制，唯一要求是不得在因果结构中引入有向环。
[14]究竟哪一种合适，取决于抽象中是否存在“变量对齐”，参见第5节。

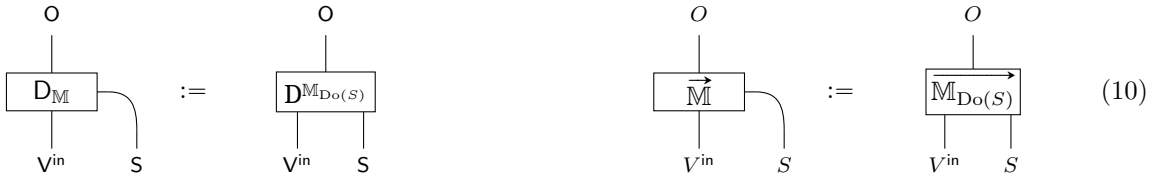
其中， x 是 X的尖锐状态。更一般地，对于 $\mathsf{S} \subseteq \mathsf{V}^{\mathsf{in}^\perp}$ 以及 S 的尖锐状态 s，Do-干预 $\mathrm{Do}(S=s)$ 会将每个 $\mathsf{X} \in \mathsf{S}$ 的机制替换为 s 在 X上的对应因子。

将 Do-干预形式化为查询很直接。给定具体变量 $\mathbb{V}$，查询集 $\mathrm{Do}(\mathbb{V})$ 具有类型 V；对于每一对子集 $\mathsf{S} \subseteq \mathsf{V}^{\mathsf{in}^\perp}$、$\mathsf{O} \subseteq \mathsf{V}^{\mathsf{out}}$ 以及 S的尖锐状态 s，都有一个查询 $\mathrm{Do}(S=s) : \mathsf{V}^{\mathsf{in}} \to O$。因此，定义在 $\mathbb{V}$ 上的因果模型 $\mathbb{M}$给出一个 $\mathrm{Do}(\mathbb{V})$ 的模型，其形式为



对于 $\mathbf{C} = \mathbf{FStoch}$ 中的因果贝叶斯网络，上述通道通常记作 $P(O \mid \mathsf{V}^{\mathsf{in}}, \mathrm{Do}(S=s))$；下文将给出一个示例。[15] 一般来说，也可以考虑查询签名 $\mathcal{I}$，其类型与 $\mathrm{Do}(\mathbb{V})$相同，但只包含其中严格子集的查询，这些查询对应于文献中常见的“干预集合” [BH19, RWB+17]。

抽象 Do 查询。可以说，Do-干预的本质是“纯粹句法上的”，与赋给 S的具体值 s 无关。受此启发，给定因果模型 $\mathbb{M}$ 和子集 $S \subseteq \mathsf{V}^{\mathsf{in}^\perp}$，我们将 $\mathbb{M}_{\mathrm{Do}(S)}$ 定义为一个具有相同变量（及输出）的因果模型，只需删除机制 c_{X}（其中 $X \in S$），从而使每个 $X \in S$ 现在都成为输入。[16] 作为一种图形约定，我们在 $\mathbf{C}$ 中将每个由此得到的输入输出图和通道表示如下。



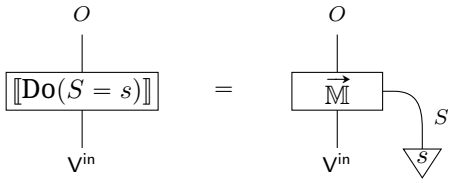
(10)

对于 $\mathbf{C}$ 中的因果贝叶斯网络，$\mathbf{C} = \mathbf{FStoch}$ 上方右侧的通道是通常记作以下形式的概率通道：

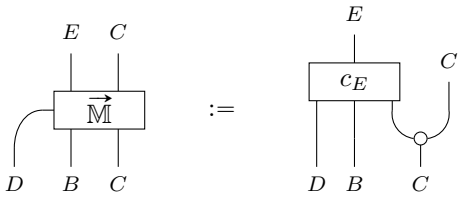
$$P(O \mid V^{\mathsf{in}}, \mathrm{Do}(S)) \tag{11}$$

形式化地，给定输入输出变量 V，我们定义由类型 V 构成的抽象 *Do* 查询 $\mathrm{Do}(\mathsf{V})$ 集合，并针对每对子集 $\mathsf{S} \subseteq \mathsf{V}^{\mathsf{in}^\perp}$、$\mathsf{O} \subseteq \mathsf{V}^{\mathsf{out}}$定义一个查询 $\mathrm{Do}(\mathsf{S}): \mathsf{V}^{\mathsf{in}} \otimes \mathsf{S} \to \mathsf{O}$。关于相应具体变量的因果模型 $\mathbb{M}$ $\mathbb{V}$ 随后给出一个将 $\mathrm{Do}(\mathsf{V})$ 视为抽象查询的模型，其中 $|-|$和 $[\![-]\!]$ 分别将每个此类查询 $\mathrm{Do}(\mathsf{S})$ 映射到式（10）的左端和右端。

请注意，Do 查询 $\mathrm{Do}(S=s)$（我们也称其为具体Do 查询）是通过组合 th e 抽象 Do 查询 $\mathrm{Do}(S)$ 具有如下尖锐状态 s。

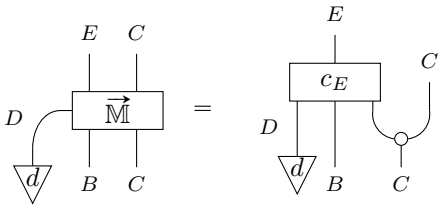


例13。 继续我们的例子，查询$\mathrm{Do}(S)$ 对于$\mathsf{S} = \{\mathsf{D}\}$ 和$\mathsf{O} = \{\mathsf{E},\mathsf{C}\}$，具有语义 ：



[15]在文献中，模型往往不包含输入，因此常见表达式为 $P(O \mid \mathrm{Do}(S=s))$。[16]因此， $\mathbb{M}_{\mathrm{Do}(S)}$ 是开放有向无环图 $G_{\mathrm{Do}(S)}$ 上的因果模型，由 G通过删除所有进入 S的边，并将 S 中的所有顶点标记为输入来定义。$\mathrm{Do}(S)$ 是因果模型之间的转移，但并非上述意义上的干预，因为它改变了输入变量的集合。进一步细节见 [LT23] ，其中称之为“开放操作”，以及 [RERS23] 中相关的“固定”操作。

对于 D的任意尖锐状态 d，具体 ***Do*** 查询 $\mathrm{Do}(D=d)$ 的语义为：

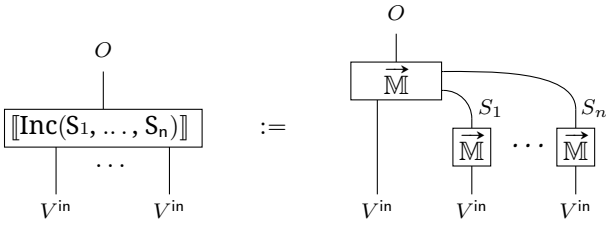


对于因果贝叶斯网络，这些分别是通常记为 $P(E,C \mid B,C,\mathrm{Do}(D))$ 和 $P(E,C \mid B,C,\mathrm{Do}(D=d))$的通道。

3.2.4 交换查询

对于人工智能中使用的因果模型（例如神经网络），[17] 只考虑将取值设为模型实际产生的那些值的干预，可能更为自然。

给定输入输出变量 V，我们将（抽象）交换查询 Inc(V) 的查询集定义为具有类型 V，并且对于每个子集 $\mathsf{O} \subseteq \mathsf{V}^{\mathsf{out}}$ 以及一组不相交子集 $\mathsf{S}_1, \ldots, \mathsf{S}_n \subseteq \mathsf{V}^{\mathsf{in}^\perp}$，有一个从 $n+1$ 份 V^{in} 到 O 的查询 $\mathsf{Inc}(S_1, \ldots, S_n)$。任何具有变量 V 的因果模型 $\mathbb{M}$ 都会产生一个将 Inc(V) 作为抽象查询的模型，其中：

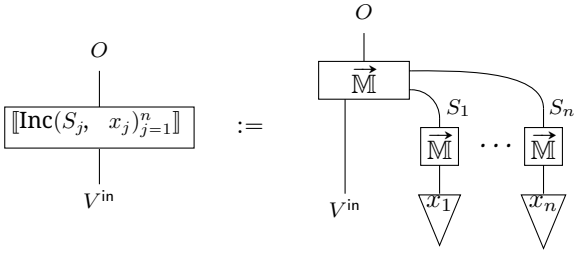


(12)

与 Do 查询一样，交换查询也有具体形式，事实上它们是普通 Do-干预 $\mathrm{Do}(S=s)$ 的特例。给定不相交子集 S_1、…、$\mathsf{S}_n \subseteq \mathsf{V}^{\mathsf{in}^\perp}$ 以及 V^{in} 的尖锐状态 x_1、…、x_n，对 $\mathbb{M}$ 的交换干预 形式如下：

$$\mathsf{Inc}(S_j, x_j)_{j=1}^n := \mathrm{Do}(S_j = \vec{\mathbb{M}} \circ x_j)_{j=1}^n$$

给定具体变量 $\mathbb{V}$，我们将带有类型 V 的查询签名记为 Inc($\mathbb{V}$)；它包含针对每种可能的交换干预以及输出集合 O 的一个查询，也称为具体交换查询。因果模型 $\mathbb{M}$ 通过对一组尖锐状态 x_1、…、x_n应用交换查询，来表示任何此类查询。



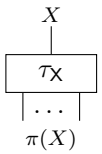
4 抽象

我们现在引入模型之间抽象的核心概念。始终固定一个(c)d-范畴$\mathbf{C}$，并用 $(\mathbb{M}, \mathcal{Q})$ 表示由查询签名 $\mathcal{Q}$ 以及$\mathbf{C}$中的模型 $\mathbb{M}$ 组成的一对，其函子

[17]任何神经网络都以纯形式化的方式构成一个因果模型；另一个问题是，这些是否是“有意义”或“可解释”的因果变量。

$[\![-]\!]_{\mathbb{M}}$: $\mathbf{Q} \to \mathbf{C}$。[18] 事实上，我们将遇到两种抽象概念，二者都将一个低层级 模型 $(\mathbb{L}, \mathcal{Q}_L)$与一个高层级 模型 $(\mathbb{H}, \mathcal{Q}_H)$联系起来，并共享以下基本要素。回想一下，在任何范畴中，态射 f 是满态射 （即满态射），如果每当 $g \circ f = h \circ f$ 时我们有 $g = h$；这是集合之间满射函数的范畴论推广。

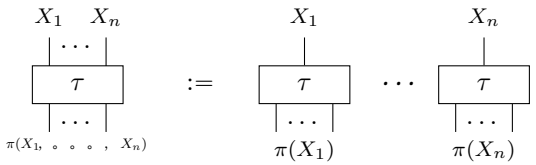
定义 14 （类型对齐）。 一个类型对齐 $(\pi, \tau)\colon (\mathbb{L}, \mathcal{Q}_L) \to (\mathbb{H}, \mathcal{Q}_H)$ 通过以下方式给出：对于 $\mathcal{Q}_H$ 的每个类型X，指定一列类型 $\pi(\mathsf{X})$，这些类型属于 $\mathcal{Q}_L$，并附带一个满态射确定性信道，其位于$\mathbf{C}$中：



(13)

其中 $X = [\![\mathsf{X}]\!]_{\mathbb{H}}$, $\pi(X) := [\![\pi(\mathsf{X})]\!]_{\mathbb{L}}$。

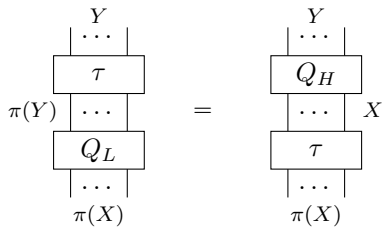
此后，在图表中，对于每个 τ_X，我们只写作 τ，并通过线标签将它们区分开来。直观而言，类型对齐构成了在一对模型之间传递的数据：它将低层级模型 $\mathbb{L}$ 的查询类型 $\mathbf{X}$) 指派给高层级模型 $\mathbb{H}$的每个查询类型 $\mathbf{X}$；反过来，也给出一个映射 τ，将前者的低层级状态映射到后者的高层级状态。我们将其扩展到任意类型列表：将 $\pi(\mathsf{X}_1、\ldots、\mathsf{X}_n)$ 定义为 $\pi(\mathsf{X}_j)$ 的串接，并将 τ 定义为幺半积：[19]



(14)

虽然对于**FStoch**中的满确定态射而言这是自动成立的，但一般而言，当任意 $X \in \mathsf{V}_H$ 的每个尖锐状态都具有形式 $\tau \circ s$ （对应于 $\pi(X)$的某个尖锐状态 s ）时，我们称 τ 是满射。

现在，在一个抽象中，我们还希望将各层级的查询本身联系起来，并且有两种主要方式可以做到这一点。在每种情况下，我们将每个高层查询Q_H: $\mathsf{X} \to \mathsf{Y}$与至少一个低层查询$\mathsf{Q}_L$: $\pi(\mathsf{X}) \to \pi(\mathsf{Y})$联系起来，使得一致性在$\mathbf{C}$中成立：

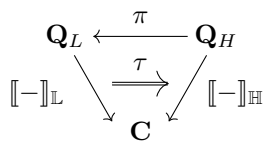


(15)

其中 $Q_H = [\![\mathsf{Q}_H]\!]_{\mathbb{H}}$, $Q_L = [\![\mathsf{Q}_L]\!]_{\mathbb{L}}$。

向下抽象。第一种此类关系涉及查询上从高层到低层层级的映射 π，并带有直接的范畴论描述。

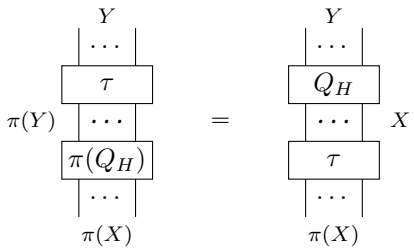
定义 15 （向下抽象）。 一个向下抽象 $(\mathbb{L},\ \mathcal{Q}_L) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H},\ \mathcal{Q}_H)$ 由函子 π 和如下形式的满自然变换 τ 给出：



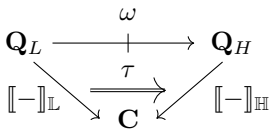
[18]虽然这就是本节明确要求的全部数据，但在实践中，我们通常会有一个包含更多结构 $\mathcal{S}$ 的组合模型 $(\mathbb{M}, \mathcal{S}, \mathcal{Q}, \mathbf{C})$，正如第 5 至 7 节中所有后续示例一样。[19]我们还要求每个这样的积态射（14）都是满态射；在 **FStoch** 这类范畴中，这一条件会自动满足。另请注意，尽管公式（14）并不要求不同的 j 对应的 $\pi(\mathsf{X}_j)$ 彼此不相交，但在我们的示例中，它们实际上总是互不相交的；不过，这并不意味着底层组合模型具有 *disjoint variable alignment*（见第 5 节），因为 $\mathcal{Q}_L$ 和 $\mathcal{Q}_H$ 可能分别只有两种类型：$_L\mathsf{in}_L$,out_H 和 $_H$，且 $\pi(\mathsf{in}_H) = \mathsf{in}_L$ 和 $\pi(\mathsf{out}_H) = \mathsf{out}_L$。

使得对于 $\mathcal{Q}_H$, π(的每个查询Q，) 都是 $\mathcal{Q}_L$的一个查询。[20]

展开定义可知，向下抽象恰好是一种类型对齐（由类型上的映射 π 和变换 τ给出），再加上一个映射 π，它将每个高层查询 $\mathsf{Q}_H : \mathsf{X} \to \mathsf{Y}$ 映射为一个低层查询 $\mathsf{Q}_L := \pi(\mathsf{Q}_H) : \pi(\mathsf{X}) \to \pi(\mathsf{Y})$。 τ的自然性无非是要求一致性 (15) 对 $\mathcal{Q}_H$中的每个查询 Q_H都成立，其中 $\mathsf{Q}_L = \pi(\mathsf{Q}_H)$。也就是说，在 $\mathbf{C}$中，我们有：



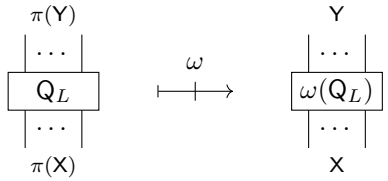
向上抽象。 然而，对于许多"具体"查询，通常会有多个低层查询与单个高层查询相关联（例如，特定的 Do-干预以及因果模型的精确变换这一概念 [RWB+17, BH19]；参见第 5 节）。为此，我们考虑第二种抽象概念，改用查询从低层到高层的偏映射 ω 。直观地说， τ 由此构成如下的自然变换。



(16)

更确切地说，我们定义如下。

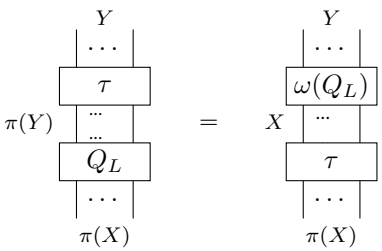
定义 16 (向上抽象**)**。 一个向上抽象 $(\mathbb{L},\ \mathcal{Q}_L)\ \overset{\tau}{\underset{\pi}{\rightleftarrows}}\ (\mathbb{H},\ \mathcal{Q}_H)$ 由一个类型对齐 $(\pi,\ \tau)$ 以及一个部分满射映射 $\omega\colon\ \mathcal{Q}_L \twoheadrightarrow \mathcal{Q}_H$ 构成，该映射作用于以下形式的查询：



(17)

使得一致性 (15) 对 $\mathsf{Q}_H = \omega(\mathsf{Q}_L)$成立。

因此，在向上抽象中，每当 $\omega(\mathsf{Q}_L)$ 有定义且类型为 $\mathsf{X} \to \mathsf{Y}$ 时，Q_L 的类型为$\pi(\mathsf{X}) \to \pi(\mathsf{Y})$，并且以下关系在 $\mathbf{C}$中成立：



(18)

这再次告诉我们，向上抽象是如示意图 (16) 所示的某种自然变换。我们在附录 F.1 的命题 78 中，通过将向上抽象与向下抽象联系起来，对此作出精确说明。

一个有用的事实是，这两种抽象概念都具有如下意义上的复合性。

[20]也就是说，是签名 $\mathcal{Q}_L$ 的，而不仅仅是范畴$\mathbf{Q}$的$_L$

命题 1 **7.** 向下抽象和向上抽象各自都在复合运算下封闭。也就是说，如果 :

$$(\mathbb{L}, \mathcal{Q}_L) \underset{\pi}{\overset{\tau(\omega)}{\rightleftarrows}} (\mathbb{M}, \mathcal{Q}_M) \underset{\pi'}{\overset{\tau'(\omega')}{\rightleftarrows}} (\mathbb{H}, \mathcal{Q}_H)$$

是向下抽象（分别为向上抽象），那么以下也是：

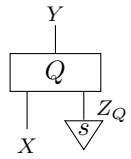
$$(\mathbb{L}, \mathcal{Q}_L) \underset{\pi \circ \pi'}{\overset{\tau' \circ \tau \ (\omega' \circ \omega)}{\rightleftarrows}} (\mathbb{H}, \mathcal{Q}_H)$$

因此，类型对齐也在复合运算下封闭。

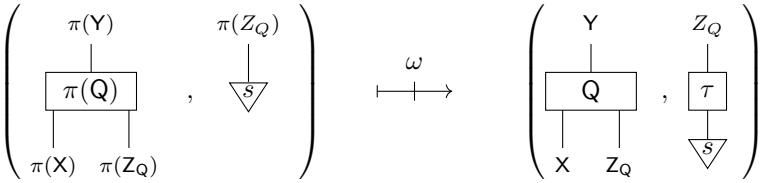
证明。 不难验证。 □

从向下抽象到向上抽象。 正如我们将在接下来的例子中反复看到的，向上抽象通常只需将向下抽象的一致性条件限制在某些输入状态上即可得到。在下一节中，我们将看到若干这样的因果向下抽象与向上抽象配对示例；它们通过同一类查询的对应抽象版本和具体版本相关联，概述见图1。

这些可以从以下一般性结果中看出。考虑一个查询模型 $(\mathbb{M}, \mathcal{Q})$，其中每个查询 Q 都附带一个选定的输入集合 $\mathsf{Z_Q}$。我们定义一个查询签名 $\mathcal{Q}^{\mathsf{Sh}}$，其类型与 $\mathcal{Q}$ 相同，并且对每个 $\mathsf{Q} \in \mathcal{Q}$ 和 $\mathsf{Z_Q}$ 的尖锐状态 s，都有一个查询 $(\mathsf{Q},\ s)$，其中 $[\![(\mathsf{Q}, s)]\!]$ 在 $\mathbf{C}$ 中定义如下：

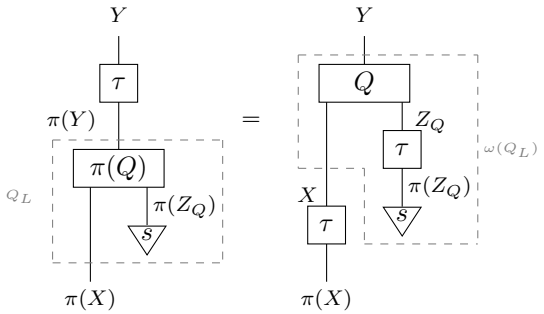


命题18. 设 $(\mathbb{L}, \mathcal{Q}_L) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H}, \mathcal{Q}_H)$ 是一个向下抽象，其中 $\mathcal{Q}_H$ 具有上述形式，且 τ 为满射。我们通过以下方式得到一个向上抽象 $(\mathbb{L}, \mathcal{Q}_L^{\mathsf{Sh}}) \rightleftarrows (\mathbb{H}, \mathcal{Q}_H^{\mathsf{Sh}})$：



其中 $\mathsf{Z}^{\pi(\mathsf{Q})} := \pi(\mathsf{Z_Q})$。

证明。 满射性由构造得出，而一致性则直接源自向下抽象的自然性，并应用于每个状态 s：



(19)

□

层级	抽象	具体	抽象类型
$\mathcal{L}^2$	Do 查询	Do-干预 （具体 Do 查询）	构造性抽象
$\mathcal{L}^2$	交换查询	交换干预 （具体交换查询 ）	交换抽象
$\mathcal{L}^3$	反事实查询	具体反事实查询	反事实抽象

图1：第5节中几种因果抽象形式的概览； $\mathcal{L}^2$ 和 $\mathcal{L}^3$ 分别表示因果层级的第二层（干预层）和第三层（反事实层）（参见例如 [Pea09, BCII22]）。

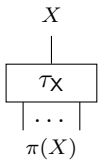
5 因果抽象

本节将介绍文献中的许多因果抽象概念如何成为我们定义的特例，即查询模型之间的自然变换。全文中，我们考虑 因果 模型 $\mathbb{L}$ 和 $\mathbb{H}$，它们属于 $\mathbf{C}$，相应的变量分别为 $\mathbf{V}^{\text{in}}_{\mathbf{L/H}}, \mathbf{V}^{\text{out}}_{\mathbf{L/H}} \subseteq \mathbf{V}^{\mathbf{L/H}}$。

在每种情况下，我们都会先以最接近文献的形式直接定义这种抽象，然后再说明，针对第 3.1 节中的因果查询作出特定选择时，它如何成为我们定义的一个特例。在若干情况下，会出现以下形式的映射。

定义19（不相交变量对齐）。 *[BH19，定义 3。19] [GIZ+23，定义 31] A* 不相交变量对齐 $(\pi,\ \tau)$: $\mathbb{V}_L \to \mathbb{V}_H$ 给出具体变量之间的对齐，定义在$\mathbf{C}$中：

- 一个集合 $ection\ of\ disjoint\ subsets\ (\pi(\mathbf{X}))_{\mathbf{X}\in\mathbf{V_H}}\ of\ \mathbf{V}_L,\ satisfying\ \pi(\mathbf{V}^{\text{in}}_{\mathbf{H}}) = \mathbf{V}^{\text{in}}_{\mathbf{L}}\ and\ \pi(\mathbf{V}^{\text{out}}_{\mathbf{H}}) \subseteq \mathbf{V}^{\text{out}}_{\mathbf{L}}$;
- 对于每个$\mathbf{X} \in \mathbf{V}_H$，在$\mathbf{C}$中都有一个确定性的满态射通道：



因此，不相交变量对齐类似于类型对齐，其中变量充当类型，并且现在要求 $\pi(\mathbf{X})$ 构成不相交子集。同样，在$\mathbf{C}$ $=$**FStoch** 中，这些条件意味着每个映射都是满射函数 τ: $\pi(X) \to X$。

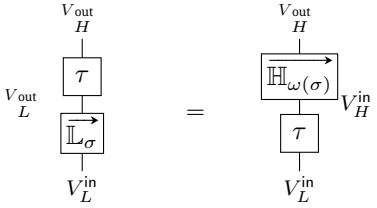
注 *20.* 对于任意具体变量 $\mathbb{V}$，记 V^* 为类型为 V 的抽象查询集合；对于 V 的每个子集 S，都有一个查询 $Q_{\mathbf{S}}$: $S \to S$，且满足 $|Q_{\mathbf{S}}|_{\mathbb{V}} = id_{\mathbf{S}}$。那么，不相交变量对齐 (π,τ): $\mathbb{V}_L \to \mathbb{V}_H$实际上是一个向下抽象 $(\mathbb{V}_L,\ V^*_L) \rightleftarrows (\mathbb{V}_H,\ V^*_H)$，满足 $\pi(V^{\text{in}}_{\mathbf{H}}) = V^{\text{in}}_{\mathbf{L}}$和 $\pi(V^{\text{out}}_{\mathbf{H}}) \subseteq V^{\text{out}}_{\mathbf{L}}$；证明见附录 A。

5.1 精确变换

在文献中，将低层级的具体干预与高层级联系起来的抽象概念，或许最早出现，也最弱，被称为模型的“精确变换”。我们将 [RWB+17，定义 3]、[BH19，定义 3.1] 和 [GIZ+23，定义 25]调整为适用于我们情形的形式如下（稍后见注 71，以进行精确比较）。

定义 21（精确变换）。 给定具有干预集合 $\mathcal{I}_L$、$\mathcal{I}_H$ 的因果模型 $\mathbb{L}$，$\mathbb{H}$，一个精确变换 $\mathbb{L} \to \mathbb{H}$ 由一对满确定性通道 τ: $V_{\text{输入}\,L} \to V_{\text{输入}\,H}$ 和 τ: $V_{\text{输出}\,L} \to V^{\text{out}}_H\ in \mathbf{C}$（我们根据其输入加以区分），以及一个满射偏函数 ω: $\mathcal{I}_L \to \mathcal{I}_H$ 给出，使得每当 ω

在 $\sigma \in \mathcal{I}_L$ 上有定义时，我们有：



(20)

回想一下，给定干预集 $\mathcal{I}_L$，查询签名 $\mathcal{I}_L^{\text{io}}$ 对每一个干预 $\sigma \in \mathcal{I}_L$都包含一个查询 σ：$\text{in}_{\mathsf{L}} \to \text{out}_{\mathsf{L}}$，对 $\mathcal{I}_H^{\text{io}}$也同样如此。

命题 22. 从 $\mathbb{L}$ 到 $\mathbb{H}$、干预集为 $\mathcal{I}_L$、$\mathcal{I}_H$ 的精确变换 $(\tau,\ \omega)$ 等价于向上抽象 $(\mathbb{L},\ \mathcal{I}_L^{\text{io}})^{\ \tau,\ \underset{\pi}{\overset{\omega}{\rightleftarrows}}}(\mathbb{H},\ \mathcal{I}_H^{\text{io}})$。

证明. 由定义直接可得；需注意，要使任何这样的向上抽象让 ω 将查询映射到查询，类型映射 π 必须由 $\pi(\text{in}_{\mathsf{H}}) := \text{in}_{\mathsf{L}}$ 以及 $\pi(\text{out}_{\mathsf{H}}) := \text{out}_{\mathsf{L}}$给出。 □

注意，这一概念并不假设不相交变量对齐。一个更强的版本可以被视为因果抽象的通用形式，仅差于要求不相交变量对齐，具体如下。对于任意因果 $\mathbb{M}$，我们将平凡干预定义为干预 $\mathbf{1}$，该干预不改变任何 $\overrightarrow{\quad}$ $\overrightarrow{\quad}$机制，因此 $\mathbb{M}_{\mathbf{1}} = \mathbb{M}$。

定义23（强因果抽象）。给定因果模型 $\mathbb{L}$，$\mathbb{H}$ 以及干预集合 $\mathcal{I}_L$、$\mathcal{I}_H$，从 $\mathbb{L}$ 到 $\mathbb{H}$ 的精确变换 $(\tau,\ \omega)$ 称为强因果抽象，如果

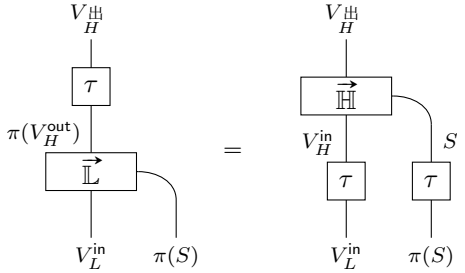
1. $\mathsf{V}^{\mathsf{in}\ \perp}_{\mathsf{H}} \subseteq \mathsf{V}^{\mathsf{out}}_{\mathsf{H}}$，即所有非输入变量都是输出变量；
2. $\mathcal{I}_H = \text{Do}(\mathbb{V}_H)$，即在高层级包含所有具体 *Do* 干预；
3. $\mathcal{I}_L$ 和 $\mathcal{I}_H$ 包含平凡干预，其中 $\omega(1_L) = 1_H$。

精确变换和强因果抽象的示例将在后续章节中给出，而对其动机及其与文献的关系的评论则留待注释 72。

5.2 构造性抽象

现在我们来谈因果抽象的一个核心、或许是最具典范性的概念。它构成了 [BH19]，所提出的最强抽象概念，尤其适用于人工智能模型，以及其他涉及将微观变量集合映射到宏观变量集合的情形。

定义 24（构造性抽象）。 *[BH19，定义 3。19]* 设 $\mathbb{L}, \mathbb{H}$ 是 $\mathbf{C}$ 中的因果模型，其具体变量分别为 $\mathbb{V}_L, \mathbb{V}_H$ 和 $\mathsf{V}^{\mathsf{in}^{\perp}}{}_{\mathsf{H}} \subseteq \mathsf{V}^{\mathsf{out}}_{\mathsf{H}}$。一个构造性抽象 $\mathbb{L} \to \mathbb{H}$ 由不相交变量对齐 (π, τ)：$\mathbb{V}_L \to \mathbb{V}_H$ 给出，使得对于所有 $\mathsf{S} \subseteq \mathsf{V}^{\mathsf{in}^{\perp}}{}_{\mathsf{H}}$，以下条件在$\mathbf{C}$中成立：



(21)

因此，构造性抽象在结构上将对 $\mathbb{H}$ 进行的（抽象）Do-干预与对 $\mathbb{L}$进行的干预关联起来。现在，我们可以用范畴论的方式将其刻画如下。

定理25。具有不相交变量对齐 (π的构造性抽象 $\mathbb{L} \to \mathbb{H}, \tau$) 等价于向下抽象 $(\mathbb{L}, \mathrm{Do}(\mathsf{V}_L)) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H}, \mathrm{Do}(\mathsf{V}_H))$，使得 $\pi(\mathsf{V}_\mathsf{H}^\mathsf{in}) = \mathsf{V}_\mathsf{L}^\mathsf{in}$。

证明。 附录 A. □

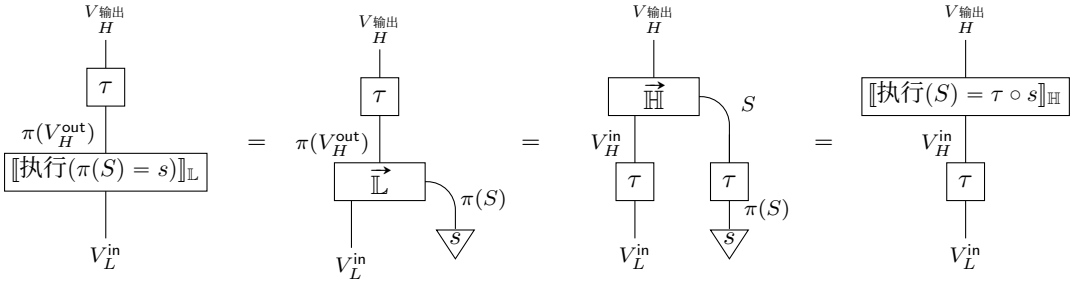
在此，我们将构造性抽象揭示为一个完全结构性的概念，即以抽象的 Do-查询来表述，而通常的表述则是带有具体值的 Do-干预。与后者的联系则可通过与输入状态的复合而直接得到。下文表明，我们也可以将构造性抽象视为 Do-干预之间的一种向上抽象，它必须将低层级 Do-干预 $\mathrm{Do}(S_L = s_L)$ 满射地映射到高层级的 $\mathrm{Do}(S_H = s_H)$。

推论26。 任何具有满射 τ 的构造性抽象 $\mathbb{L} \to \mathbb{H}$ 都定义了如下形式的向上抽象：$(\mathbb{L}, \mathrm{Do}(\mathbb{V}_L)) \overset{\tau}{\underset{\pi}{\rightleftarrows}}{}^{\omega} (\mathbb{H}, \mathrm{Do}(\mathbb{V}_H))$，经由：

$$\omega(\mathrm{Do}(\pi(S) = s))) := \mathrm{Do}(S = \tau \circ s)$$

因此，它也定义了 Do-干预的一种精确变换，即形式为 $(\mathbb{L}, \mathrm{Do}(\mathbb{V}_L)^{\mathsf{io}}) \overset{\tau}{\underset{\pi}{\rightleftarrows}}{}^{\omega} (\mathbb{H}, \mathrm{Do}(\mathbb{V}_H)^{\mathsf{io}})$ 的向上抽象，其中每个 $\mathrm{Do}(\mathbb{V})^{\mathsf{io}}$ 都被理解为仅含两种类型的签名 $\mathcal{I}^{\mathsf{io}}$。事实上，它定义了一种强因果抽象（定义 23）。

证明。 这种向上抽象的一致性条件恰好是：

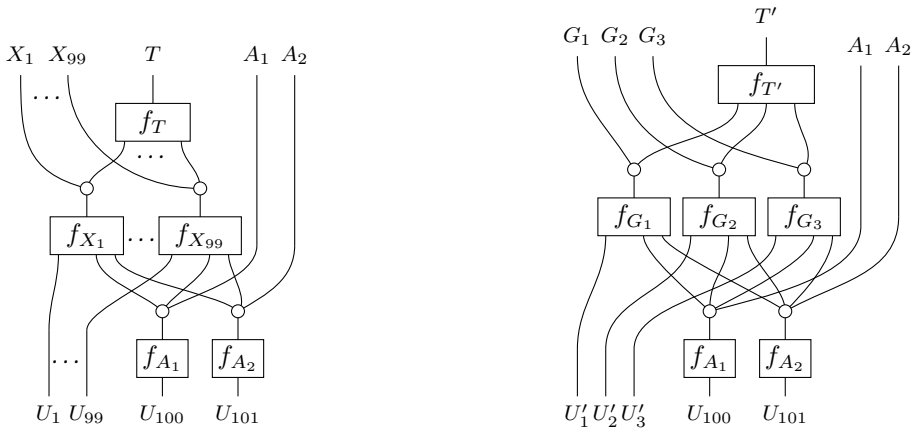


(22)

或者，这构成了命题18的一个特例，其中 $\mathcal{Q}_H = \mathrm{Do}(\mathsf{V}_H)$ 和 $\mathcal{Q}_L = \mathrm{Do}(\mathsf{V}_L)$；对于 $\mathsf{Q} = \mathrm{Do}(\mathsf{S})$，我们取 $\mathsf{Z}_\mathsf{Q} = \mathsf{S}$。由此可将式 (22) 识别为式 (19) 的一个实例。查询类型仅为输入、输出的情形显然成立。

□

例 27。 下面这个构造性抽象的简单示例出自 *[BH19]*.[21] 考虑左下方和右下方分别展示的 **FStoch** 中的开放确定性因果模型 $\mathbb{L}$ 和 $\mathbb{H}$。



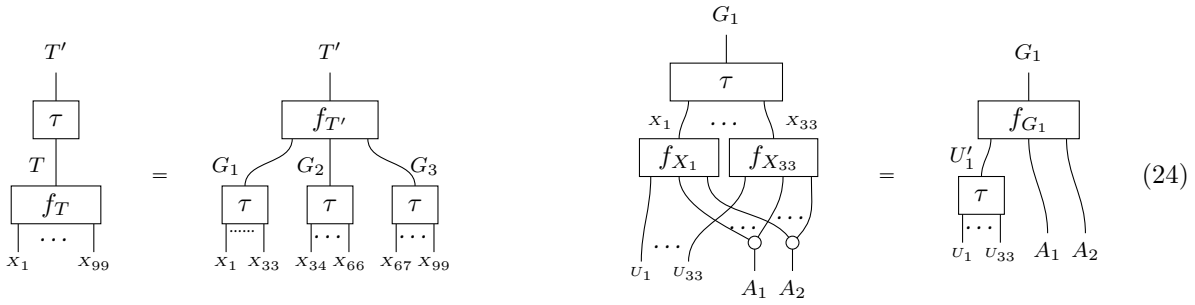
(23)

这里的 $X_1, ..., X_{99}$ 是二元变量，取值为 $\{0, 1\}$，表示 *99* 名个人对某项请愿书的投票（*"1"*表示"赞成"），T 是支持请愿书的总人数（即 $f_T = \sum_i X_i$），而 A_1, A_2 是

[21]最初来自 [R WB+17] 尽管在那里是以体现精确变换的受限干预集进行讨论的 s.

一些离散变量表示请愿书宣传活动可以选择投放的两种可能广告； $\{f_{X_i}\}_{i=1}^{99}$、 f_{A_1} 和 f_{A_2} 是相应类型的一些函数，其中 $U_1, ..., U_{101}$ 是被视为（外生）输入的背景变量。高层模型源于将选民划分为三个不同的群体 G_1、G_2 和 G_3。 G_1 取值于 $[1、 33]$，其中 $U_1' = \prod_{i=1}^{33} U_i$ 和 $f_{G_1}(a_1、 a_2、 u_1') = \sum_{i=1}^{33} f_{X_i}(a_1、 a_2、 (u_1')_i)$，对 G_2、U_2'、 f_{G_2} 和 G_3、 U_3'、 f_{G_3} 亦然；最后， T' 是二元的，表示请愿书是否被接受，其中 $f_{T'}(g_1、 g_2、 g_3)$ 在 $g_1 + g_2 + g_3 > 49$ 时取值为一，否则为零。

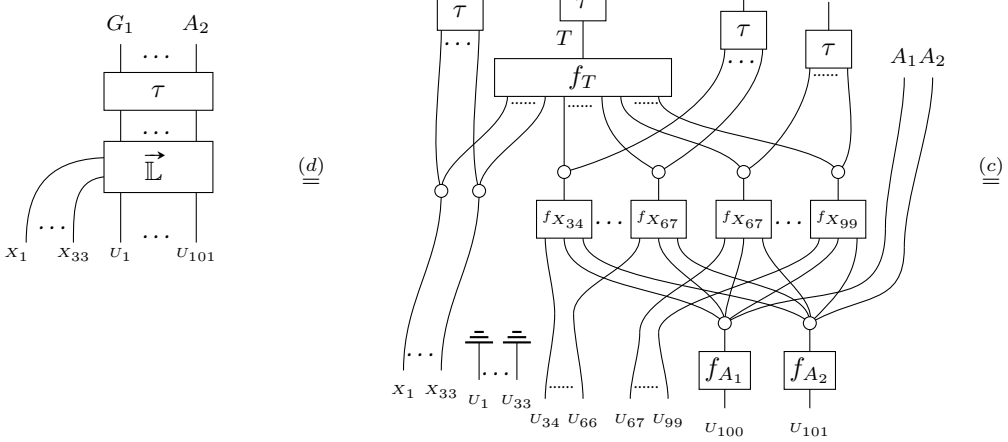
两个模型之间的关系可以用不相交变量对齐来重新表述，其中 $\pi(G_1) = \{X_1, ..., X_{33}\}$, $\pi(U_1') = \{U_1, ..., U_{33}\}$， G_2， U_2' 和 G_3， U_3' 也同理，并且 $\pi(T') = \{T\}$, $\pi(A_1) = \{A_1\}$,$\pi(U_{100}) = \{U_{100}\}$, $\pi(A_2) = \{A_2\}$, $\pi(U_{101}) = \{U_{101}\}$；映射 τ 的定义是：若 $t > 49$ ，则令 $\tau_{T'}(t) = 1$ ，否则令0 $\tau_{G_1} = \sum_{i=1}^{33} X_i$， G_2 和 G_3 也同理；并令 τ 在（外生）输入变量及 A_1， A_2 上为恒等映射。于是有：

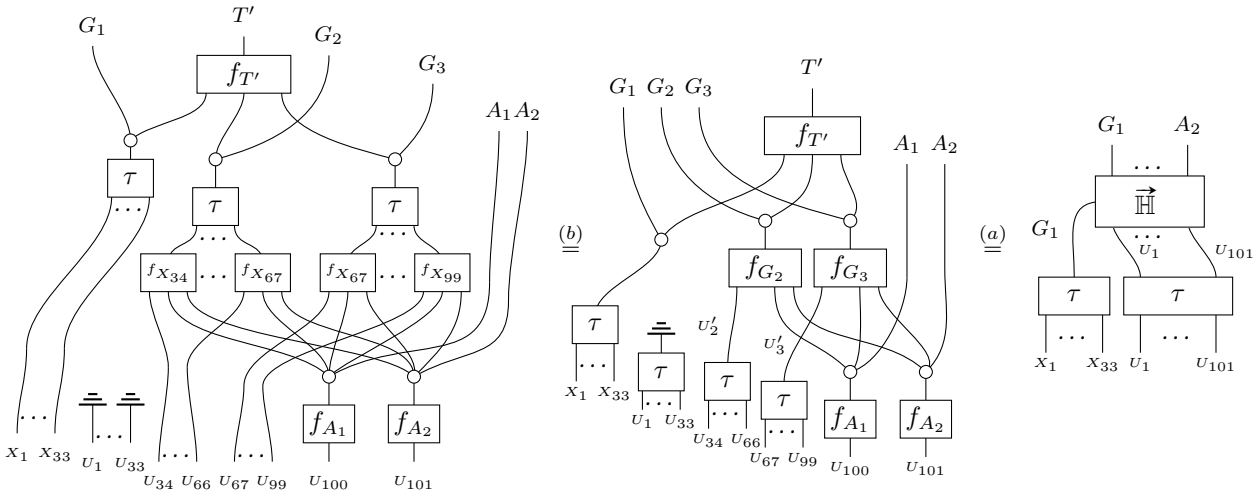


(24)

对于 f_{G_2} 和 f_{G_3}，也可以类似地处理。

可以检查，这些数据确实定义了一个构造性抽象 $\mathbb{L} \to \mathbb{H}$。作为说明性示例，我们验证对 $S = \{G_1\}$ 进行抽象 *Do*-干预时的一致性；其他情况也可用类似方式处理。此处从左到右的顺序与公式 (21) 相匹配，但实际计算从右到左最为简便：***(a)*** 和 ***(d)*** 根据定义成立；***(b)*** 使用了类似于公式(24)***for*** f_{G_2} ***and*** f_{G_3}右端的等式；***(c)*** 将确定性态射 τ 的三个实例穿过复制映射，然后利用公式 (24)左端进行改写。

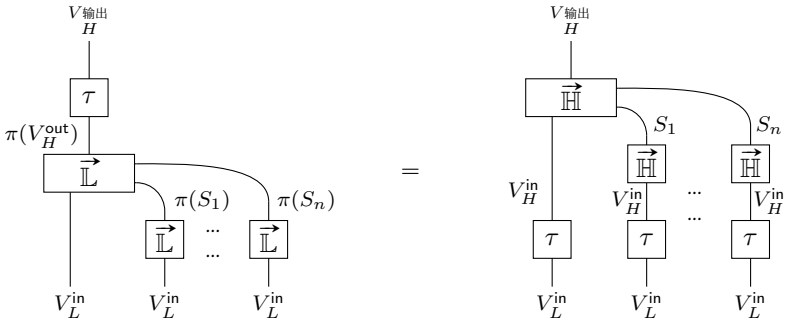




5.3 交换抽象

在实践中，可能无法针对任意输入验证构造性因果抽象条件，而只能验证那些由低层模型本身产生的输入（或其子集）。这引出了以下一种较弱的概念，它改为基于交换查询 [GLIP21, GRP20]。

定义 28（交换抽象）。一个交换抽象 $\mathbb{L} \to \mathbb{H}$ 由一个不相交变量对齐 $(\pi、\ \tau)$：$\mathbb{V}_L \to \mathbb{V}_H$给出，使得对于所有不相交子集$\mathsf{S}_1$、.....、$\mathsf{S}_n \subseteq \mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{H}$，以下条件在$\mathbf{C}$中成立：



(25)

上述内容对应于交换查询 Inc(V) 之间的一致性条件，我们可以将其形式化如下。

命题 29。具有不相交变量对齐 $(\pi$的$\tau)$ 交换抽象 $\mathbb{L} \to \mathbb{H}$ 等价于一个向下抽象 $(\mathbb{L},\ \mathsf{Inc}(\mathsf{V}_L)) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H},\ \mathsf{Inc}(\mathsf{V}_H))$，满足 $\pi(\mathsf{V}^{\mathsf{in}}_{\mathsf{H}}) = \mathsf{V}^{\mathsf{in}}_{\mathsf{L}}$、$\pi(\mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{H}) \subseteq \mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{L}$，且：

$$\pi\big(\mathsf{Inc}(\mathsf{S}_1, \ldots, \mathsf{S}_n) : \mathsf{V}^{\mathsf{in}}_{\mathsf{H}} \to \mathsf{X}\big) \;=\; \mathsf{Inc}(\pi(\mathsf{S}_1), \ldots, \pi(\mathsf{S}_n)) : \mathsf{V}^{\mathsf{in}}_{\mathsf{L}} \to \pi(\mathsf{X}) \tag{26}$$

证明。附录 A。 □

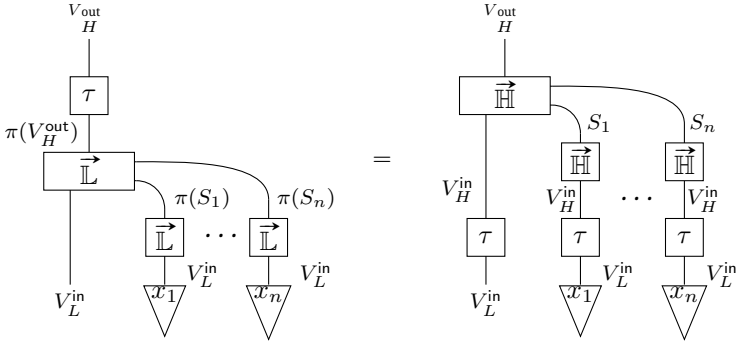
正如构造性抽象与具体 Do 干预相关联，交换抽象也与具体交换干预相关联。向上抽象 $(\mathbb{L}、\mathsf{Inc}(\mathbb{V}_L)) \rightleftarrows (\mathbb{H}、\mathsf{Inc}(\mathbb{V}_H))$ 必须以一致的方式将低层交换干预满射映射到高层级干预，而交换抽象提供了一种结构化的方法来实现这一点。

推论 30. 任何带有满射 τ 的交换抽象 $\mathbb{L} \to \mathbb{H}$ 都定义了一种形式为 $(\mathbb{L},\ \mathsf{Inc}(\mathbb{V}_L))\ ^{\tau},\ \underset{\pi}{\overset{\omega}{\rightleftarrows}}\ (\mathbb{H},\ \mathsf{Inc}(\mathbb{V}_H))$ 的向上抽象，具体如下：

$$\omega(\mathsf{Inc}(\pi(S_j), x_j)_{j=1}^n) := \mathsf{Inc}(S_j, \tau \circ x_j)_{j=1}^n$$

与构造性抽象一样，这种 u-抽象也给出了具体（交换）干预 $(\mathbb{L}、\mathrm{Inc}(\mathbb{V}_L)^{\mathsf{io}})^{\tau,\ \overset{\omega}{\underset{\pi}{\rightleftarrows}}}(\mathbb{H}、\mathrm{Inc}(\mathbb{V}_H)^{\mathsf{io}})$ 的精确变换。

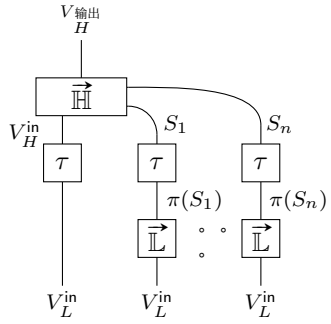
证明。 针对这类 u-抽象的查询一致性条件，直接化为以下陈述：



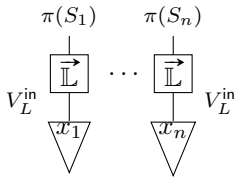
(27)

这可直接由 (25) 得出。或者，应用命题18，其中 $\mathcal{Q}_H = \mathrm{Inc}(\mathsf{V}_H)$, $\mathcal{Q}_L = \mathrm{Inc}(\mathsf{V}_L)$。对于 $\mathsf{Q} = \mathrm{Inc}(\mathsf{S}_1, \ldots, \mathsf{S}_n)$ 我们取 $Z_Q = S_1 \otimes \cdots \otimes S_n$。 □

注 31。这推 导出交换抽象关系意味着我们可以将(25) 的右端改写为：[2] [2]



因此，交换抽象恰好表示构造性抽象成立，但仅限于具有以下形式的 $\pi(S)$ 输入状态，且$\mathsf{S} = \mathsf{S}_1 \cup \cdots \cup \mathsf{S}_n$ 彼此不相交。

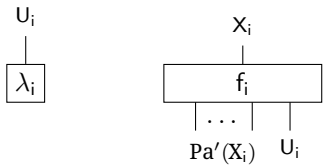


(28)

5.4 反事实抽象

现在让我们 首先指出定义反事实所依据的因果模型类型 .

定义32. 马尔可夫范畴$\mathbf{C}$ 中的函数因果模型（FCM） $\mathbb{M}$ 是一种因果模型，其变量划分为$\mathsf{V} = \mathsf{E} \cup \mathsf{U}$，其中$\mathsf{V}^{\text{out}} = \mathsf{E}$，$\mathsf{V}^{\text{in}} = \varnothing$。外生变量$\mathsf{U} = \{\mathsf{U}_i\}_{i=1}^n$满足：每个$\mathsf{U}_i$ 都没有父节点，其机制如左下方所示；而内生变量 $\mathsf{E} = \{\mathsf{X}_i\}_{i=1}^n$ 具有右下方所示的形式：



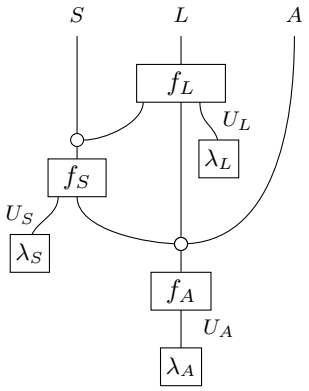
(29)

其中$\mathrm{Pa}'(\mathsf{X}_i) \subseteq \mathsf{E}$ 且 $f_i = [\![\mathsf{f}_i]\!]_{\mathbb{M}}$ 是确定性的（2）。

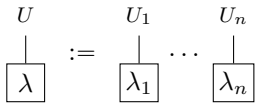
[22]该重写使用了（25）的特殊情形，其中 $\mathsf{S}_1 = \cdots = \mathsf{S}_n = \varnothing$。

例 33。 在 $\mathbf{C} = \mathbf{FStoch}$ 中的 *FCM*，恰好就是通常意义上的结构因果模型（SCM）。它由有限集 X_i，U_i 以及对每个 i 的一个分布 λ_i（定义在 U_i 上）和一个函数 $f_i\colon \mathrm{Pa}'(X_i) \times U_i \to X_i$组成，其中 $\mathrm{Pa}'(X_i) \subseteq E = \{X_1, \ldots, X_n\}$。

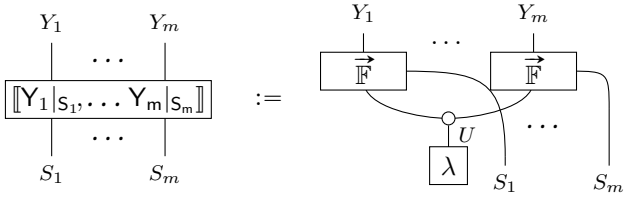
示例 34. 下图展示了一个包含$\mathsf{E} = \{\mathsf{S}、\mathsf{L}、\mathsf{A}\}$ 以及$\mathsf{U} = \{\mathsf{U_S}、\mathsf{U_L}、\mathsf{U_A}\}$的 *FCM* 网络图。



与之密切相关的是以下概念。我们所说的确定性因果模型 $\mathbb{F}$，是指其机制 c_X 对所有 $X \in \mathbf{V}^{\mathsf{in}^\perp}$ 都是确定性的因果模型。[23] 形式上，我们可以将任意 FCM 视为两个开放因果模型的复合[24]$\mathbb{M} = \mathbb{F} \circ \mathbb{U}$：$\mathbb{F}$ 是一个定义在 E 上的确定性模型，U以 U作为输入，并包含所有确定性机制 f_i；$\mathbb{U}$ 是一个仅包含 U 及其机制、对应于该联合状态的模型：



反事实查询。 现在，我们来考虑专为 FCM 定义的查询，遵循 [XB24]。给定抽象变量 E，定义签名 $\mathcal{L}^3(\mathbf{E})$：反事实查询 的类型为 E，并且对于$j = 1$的每一种子集选择，都有一个查询 $\mathsf{Y_1}|_{\mathsf{S_1}}, \ldots, \mathsf{Y_m}|_{\mathsf{S_m}}$，其输入为 $(S_j)_{j=1}^m$，输出为 $(Y_j)_{j=1}^m$，其中 S_j、$Y_j \subseteq$分别是 E 的子集。任意定义在 E、U 上的 FCM $\mathbb{M} = \mathbb{F} \circ \mathbb{U}$ 都会通过以下方式生成这些查询的模型：

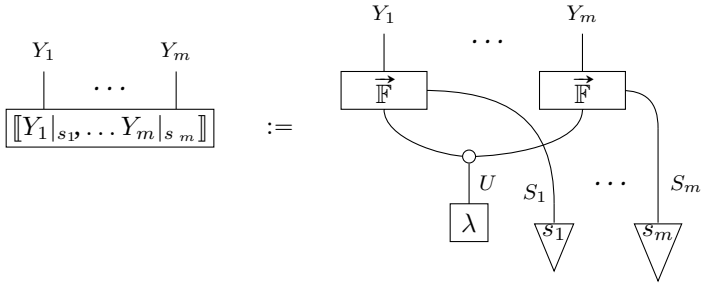


(30)

同样，给定 $\mathbb{E}$ 中的具体变量 $\mathbf{C}$，我们将具体反事实查询 的签名 $\mathcal{L}^3(\mathbb{E})$ 定义为：类型为 E 的查询 $Y_1|_{s_1}$、…、$Y_m|_{s_m}$，其输出为 $\mathsf{Y_1}$、…、$\mathsf{Y_m}$，且没有输入；对每一种子集的选择

[23]请注意，确定性因果模型并不要求变量采用某种特定的划分，因此，根节点变量要么是输入变量，要么具有一个尖锐状态（点分布）作为其“机制”。[24]关于开放因果模型之间复合这一概念的详细信息，请参见 [LT23，第 5 节。2] 复合时，会合并相应的机制集合，并识别复合所涉及的相应输入变量和输出变量。

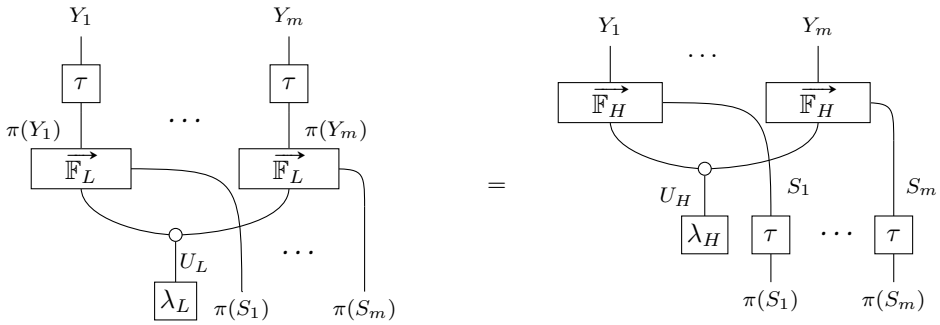
$S_j, Y_j \subseteq \mathsf{E}$ 和 sh arp 状态 s_j 属于 S_j，对于 $j = 1$、…、m，由任意一个定义在 E 上的 FCM $\mathbb{M} = \mathbb{F} \circ \mathbb{U}$ 建模为：



具体反事实查询与反事实密切相关。从本质上说，反事实由具体反事实查询所产生的条件概率分布给出；详情请参阅 [LT23]。[26]

反事实抽 象。 现在，我们可以介绍一个相关的抽象概念，它主要源自 [XB24].

定义 35 (反事实抽象**)。** 设 $\mathbb{L}, \mathbb{H}$ 是具有具体内生变量 $\mathbb{E}_L, \mathbb{E}_H$ 的 *FCM*。一个反事实抽象 $\mathbb{L} \to \mathbb{H}$ 由一个不相交变量对齐给出[27] (π, τ): $\mathbb{E}_L \to \mathbb{E}_H$，使得对于所有反事实查询 $Y_1|_{S_1}, \ldots, Y_m|_{S_m} \in \mathcal{L}^3(\mathsf{E_H})$，以下内容在 $\mathbf{C}$ 中成立：



(31)

至此，我们可以立即认出上述内容是一个自然性条件。

命题36。 反事实抽象 $\mathbb{L} \to \mathbb{H}$ 等价于如下形式的向下抽象：$(\mathbb{L}, \mathcal{L}^3(\mathsf{E_L})) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H}, \mathcal{L}^3(\mathsf{E_H}))$，其中：

$$\pi(Y_1|_{S_1}, \ldots, Y_m|_{S_m}) := \pi(Y_1)|_{\pi(S_1)}, \ldots, \pi(Y_m)|_{\pi(S_m)} \tag{32}$$

证明。 给定一个反事实抽象，按通常方式将 π, τ 扩展为变量的乘积，并通过式（32）定义一个作用于查询的映射。于是，式（31）恰好表明一致性（自然性）成立。反过来，给定这样一个向下抽象，一致性可确保式（31）成立。根据定义，π 将变量子集映射为变量子集，因此 (π, τ) 构成一个不相交变量对齐。 □

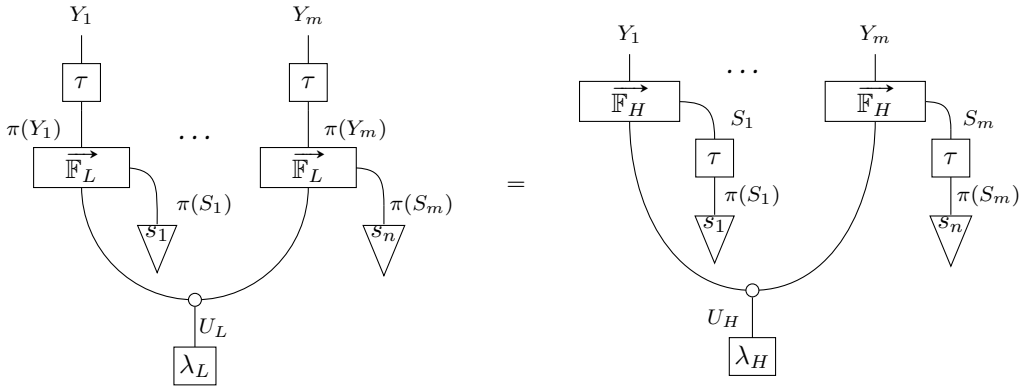
同样，也存在一种相应形式的具体向上抽象，此处适用于具体反事实查询 $\mathcal{L}^3(\mathbb{E})$。

推论37. 任何具有满射 τ 的反事实抽象 $\mathbb{L} \to \mathbb{H}$ 都确定一个向上抽象$(\mathbb{L}, \mathcal{L}^3(\mathbb{E}_L)) \rightleftarrows (\mathbb{H}, \mathcal{L}^3(\mathbb{E}_H))$，其方式如下：

$$\omega(\pi(Y_1)|_{s_1}, \ldots, \pi(Y_m)|_{s_m}) := Y_1|_{\tau \circ s_1}, \ldots, Y_m|_{\tau \circ s_m} \tag{33}$$

[25]具体变量带有指定的输入和输出子集；此处我们要求 $\mathbf{E}^{\text{in}} = \varnothing$，因为反事实查询的类型仅涵盖被视为内生的变量（在任何能生成这些查询之模型的 FCM 中）。[26]这也包括 $Y_1|_{s_1}$、…、$Y_n|_{s_n}$ 与“反事实表达式的合取”等术语之间的关系。[27]注意，对于 FCM 的内生变量，我们有 $E = \mathbf{V}^{\text{out}} = \mathbf{V}^{\text{in}^\perp}$，因此 $\pi(\mathbf{X})$ 必定只是一些不相交子集。

证明。 这种向上抽象的一致性条件为:



(34)

这可由 (31) 得出。或者，使用 $\mathcal{Q}_H = \mathcal{L}^3(E_H)$, $\mathcal{Q}_L = \mathcal{L}^3(E_L)$ 应用命题18。对于 (30) 中的每个此类查询 Q，ZQ 是 $S_1 \otimes \cdots \otimes S_n$。 □

与文献中的其他概念一样，反事实抽象此前已被引入为具体向上抽象 [XB24]。我们将关于 $\mathbb{E}$ 的一组具体反事实查询 定义为一组查询 $\mathcal{Q} \subseteq \mathcal{L}^3(\mathbb{E})$ 的子集，其类型由变量 E 给出。给定关于 $\mathbb{E}_H$ 的具体反事实查询 $\mathcal{Q}$，以及不相交变量对齐 (π, τ): $\mathbb{E}_L \to \mathbb{E}_H$，若对于 $\mathcal{Q}$ 中的所有查询 $Y_1|_{\tau \circ s_1}, \ldots, Y_m|_{\tau \circ s_m}$，式 (34) 均成立，则称 $\mathbb{L}, \mathbb{H}$ 满足 $\mathcal{Q}$-τ一致性 [XB24，定义 7]。由此立即得到以下结论。

推论38。 因果模型 $\mathbb{L}, \mathbb{H}$ 满足 $\mathcal{Q}$-τ 一致性，当且仅当数据通过 (33) 确定了一个向上抽象$(\mathbb{L}, \mathcal{L}^3(\mathbb{E}_L))^{\tau}$, $\overset{\omega}{\underset{\pi}{\rightleftarrows}} (\mathbb{H}, \mathcal{Q}_H)$，且对 $\mathcal{Q}_H$ 中的所有查询 $Y_1|_{\tau \circ s_1}, \ldots, Y_m|_{\tau \circ s_m}$ 均成立。

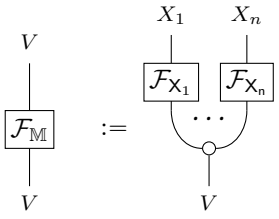
5.5 分布式因果抽象

虽然诸如构造性抽象之类的抽象概念是“局部的”，因为它们通过不相交变量对齐，将高层级的变量 V 与低层级的不相交子集 π(V) 相关联，但有人认为，不具备这一特征的更一般类型的抽象——我们可以称之为“分布式”抽象——可能更为实用，例如在机器学习 [GWP+24、GIZ+23]中。事实上，第5.1节中的精确变换和强因果抽象通常允许这种“分布式性”。在本节中，我们讨论一类特定的分布式抽象，其灵感来自 [GWP+24、GIZ+23]，并利用以下方式来定义因果模型。

5.5.1 模型归纳

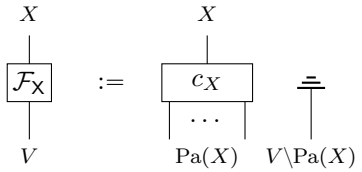
在本节中， $\mathbb{M}$ 始终指定义在变量 $\mathbb{V}$ 上的确定性开放因果模型，且 $\mathbf{V}^{\text{out}} = \mathbf{V}^{\text{in}^{\perp}}$，使得每个变量 X 在C中至少具有一个规范化状态。

定义39 （从确定性通道构建模型，反之亦然）。 我们将并行机制信道 $\mathcal{F}_{\mathbb{M}}$: $V \to V$ 定义为:



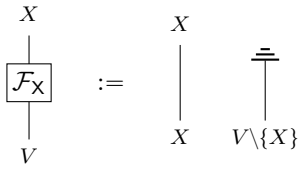
(35)

其中 $\mathbf{V} = \mathbf{X}_1, \ldots, \mathbf{X}_n$。此处，对于每个非输入变量 $\mathbf{X}$，我们设定



(36)

对于每个输入变量 $\mathbf{X}$，我们定义：



(37)

反过来，设 $\mathcal{F}\colon\ V \to V$ 是 $\underline{\mathbf{C}}$ 中的一个确定性信道，并将 $\mathcal{F}_X$ 记为 X上的每个边缘分布。我们如下定义签名 $\mathcal{S}_{\mathcal{F}}$ 和模型 $\mathcal{F}$ 。当$\mathbf{X}$ 满足(37)时，我们将其声明为输入变量。否则，我们将$\mathrm{Pa}(\mathbf{X})$ 定义为变量的最小子集，并将 c_X 定义为相应的信道，使得 $\mathcal{F}_X$按(36)中的方式分解；根据 *[LT23，引理 112]*，这样的子集和信道是存在的。接着，我们以该表示纳入生成元 $c_{\mathbf{X}}\colon \mathrm{Pa}(\mathbf{X}) \to \mathbf{X}$ 。当信道 $\mathcal{F}$ 无环 时，我们称其为无环的，即 $\mathcal{S}_{\mathcal{F}}$ 定义了一个有效的因果签名，也就是说，这些父集$\mathrm{Pa}(X)$ 在$\mathbf{V}$ 上诱导出一个无环有向图。

容易看出，如果 $\mathbb{M}$ 作为开放确定性因果模型是忠实的[28] ，那么 $\mathcal{F}_{\mathbb{M}}$ 是无环的，且$\mathcal{F}_{\mathbb{M}} = \mathbb{M}$。上述构造使我们也可以反过来从 V上的任意此类无环信道$\mathcal{F}$定义因果模型 $\mathcal{F}$，其中 $\mathcal{F}_{\mathcal{F}} = \mathcal{F}$。

连接模型的输入—输出视角与并行机制视角的一个有用事实如下。

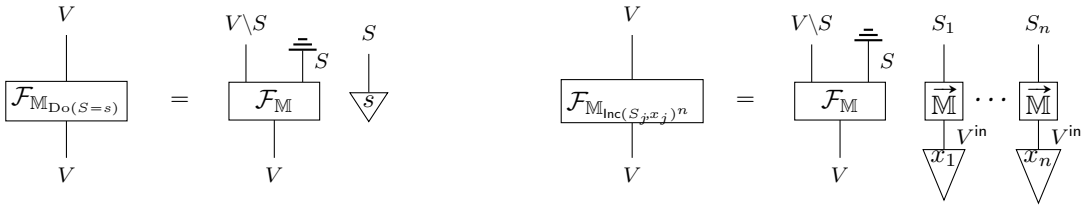
引理 40。 对于任意尖锐状态 $v = (i,\ o)$ 在 $V_{\text{中}} \otimes V_{\text{的}^{\perp}}$，我们有



(38)

证明。 取边缘分布后，后者成立当且仅当 $v_X = c_X \circ v|_{\mathrm{Pa}}(X)$ 对于所有变量 $\mathbf{X} \in \mathbf{V}^{\mathsf{in}\perp}$均成立。通过沿变量的最长祖先链进行$\vec{}$归纳法，可以看出这等价于 $o = \mathbb{M} \circ i$。 □

例 41。 直接验证可知，*Do*-干预和交换干预会如下改变 $\mathcal{F}_{\mathbb{M}}$。



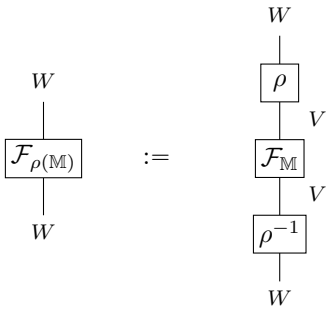
(39)

现在，我们可以利用上述构造来定义如下内容。

定义 42 （模型归纳）。设 $\mathbb{M}$ 是一个以变量 $\mathbb{V}$ 为基础的开放确定性因果模型。设 $\mathbb{W}$ 是另一组具体变量，且 ρ: $V \simeq W$ 是一个确定性同构。当通道 ρ 满足 $\mathbb{M}$

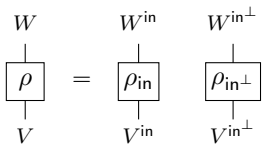
[28]从 [LT23]的机制忠实性来看： $\mathbb{M}$ 的父集 $Pa(X)$ 已经是满足式（36）的最小父集，也就是说， X 通过 c_X 依赖于其每一个父变量。

满足以下条件时：



(40)

是无环的，输入为 W^{in},[29] ，此外：



(41)

for 同构 ρ_{in}, $\rho_{\mathsf{in}^\perp}$。在这种情况下，我们定义其诱导的 因果模型 $\rho(\mathbb{M}) := \mathcal{F}_{\rho(\mathbb{M}}$ $\overline{\qquad}$ $)$.

类似地，给定在 $\mathbb{M}$ 上的干预集 $\mathcal{I}$，若 ρ 满足 $(\mathbb{M},\mathcal{I})$，即对于每个 $\sigma \in \mathcal{I}$ 都满足 $\mathbb{M}_\sigma$，我们就称其为满足条件。在这种情况下，对于每个 $\sigma \in \mathcal{I}$，我们通过 $\rho(\mathbb{M})_{\rho(\sigma)} := \rho(\mathbb{M}_\sigma)$ 定义在 $\rho(\mathbb{M})$ 上的诱导干预 $\rho(\sigma)$。我们用 $\rho(\mathcal{I})$ 表示 $\rho(\mathbb{M})$ 上此类干预的集合。

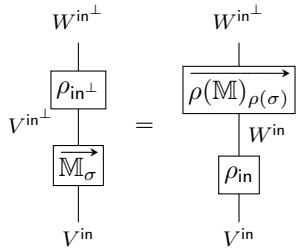
现在我们来看看，这种模型归纳确实会产生向上抽象的一个特殊情形。回想一下，在 $\mathbb{M}$ 上的任意干预 $\mathcal{I}$ 都构成一个查询签名 $\mathcal{I}^{\mathsf{io}}$，其类型为 in、out，并且对每个 $\sigma \in \mathcal{I}$ 都有一个查询 $\sigma\colon \mathsf{in} \to \mathsf{out}$。本节中，我们总是假设每个这样的集合 $\mathcal{I}$ 都包含平凡干预 1，它对应于 $\mathbb{M}$ 本身。回想一下，如命题 22 的证明所示，对于相对于查询签名 $\mathcal{I}^{\mathsf{io}}$ 的向上抽象，类型上的映射 π 只有一种可能（$\pi(\mathsf{in}_\mathsf{H})=\mathsf{in}_\mathsf{L}$ 且 $\pi(\mathsf{out}_\mathsf{H})=\mathsf{out}_\mathsf{L}$）。因此，本节余下部分将省略相应的 π，并仅用 (ω,τ) 表示向上抽象的数据。

最后，我们假设语义范畴**C** 具有足够多的状态，这意味着只要我们有 $f \circ x = g \circ x$ 对于所有尖锐状态 x，我们都有 $f = g$；这对**FStoch**成立。

命题 43. 设 $\mathbb{M}$是定义在 $\mathbb{V}$上的开开放确定性因果模型，其干预集为 $\mathcal{I}$。假设 $\rho\colon V \to W$是一个保持 $(\mathbb{M},\mathcal{I})$的确定性同构。那么

$$(\mathbb{M},\mathcal{I}^{\mathsf{io}}) \underset{\longleftarrow}{\rho,\varrho} (\rho(\mathbb{M}),\rho(\mathcal{I})^{\mathsf{io}})$$

定义了一个（双射的）向上抽象。明确地说，对于所有 $\sigma \in \mathcal{I}$，我们都有：



(42)

特别地，$\rho_{\mathsf{in}^\perp} \circ \overrightarrow{\mathbb{M}} = \overrightarrow{\rho(\mathbb{M})} \circ \rho_{\mathsf{in}}$。

证明。附录 A。 □

[29]回顾一下，作为具体变量，$\mathbb{W}$ 带有子集 W_{in}，这里的条件是，通过 (37) 定义的输入与该集合 W^{in}一致。

5.5.2 组合归纳法与抽象

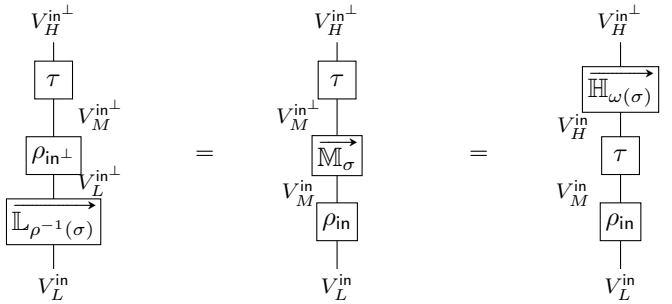
鉴于向上抽象在复合下封闭，上文讨论的模型归纳概念自然导出一种构造：考虑一个关于 $\mathbb{V}_L$ 的模型 $\mathbb{L}$ 和一个确定性同构，它会导出一个“中间”模型，而该模型又被一个高层模型 $\mathbb{H}$ 抽象。所得的复合可以在 $\mathbb{L}$ 和 $\mathbb{H}$ 之间建立一种 “ 分布式 ” 抽象关系。具体而言，假设我们有：

- 一个位于 $\mathbb{L} = \mathcal{F}_{\mathbb{L}}$ 上的“低层级”模型 $\mathbb{V}_L$；
- 一个确定性同构 ρ： $V_L \to V_M$ ，它保持 $\mathbb{L}$；
- 一个作用于 $\mathcal{I}_M$ 上的干预集 $\mathbb{M}$： $= \rho(\mathbb{L})$ ，由 ρ^{-1}保持；
- 一个向上抽象 $(\mathbb{M},\ \mathcal{I}_{\mathsf{io}_M})\ ^{\tau,}_{}\ ^{\omega}_{\rightleftarrows}\ (\mathbb{H},\ \mathcal{I}_{\mathsf{io}_H})$，指向一个位于 $\mathbb{H}$上的模型 $\mathbb{V}_H$。

然后定义 $\mathcal{I}_L$： 在 $\mathbb{L}$ 上的 $= \rho^{-1}(\mathcal{I}_M)$ ，我们得到一个向上抽象序列：

$$(\mathbb{L}, \mathcal{I}_L^{\mathsf{io}}) \xrightarrow[\leftarrow]{\rho,\rho} (\mathbb{M}, \mathcal{I}_M^{\mathsf{io}}) \xrightarrow[\leftarrow]{\tau,\omega} (\mathbb{H}, \mathcal{I}_H^{\mathsf{io}}) \tag{43}$$

由此复合得到整体向上抽象 $(\tau \circ \rho, \omega \circ \rho)$ ，从 $\mathbb{L}$ 到 $\mathbb{H}$。具体而言，对于 ω 有定义的每个干预 $\sigma \in \mathcal{I}_M$ ，我们都有：



(44)

当向上抽象 (τ, ω) 源自构造性抽象或交换抽象时，这种构造尤其值得关注。由此得到一种整体概念，我们接下来将对此进行讨论；这一概念在 [GWP+24、GIZ+23],[30] 的研究中已有隐含涉及。

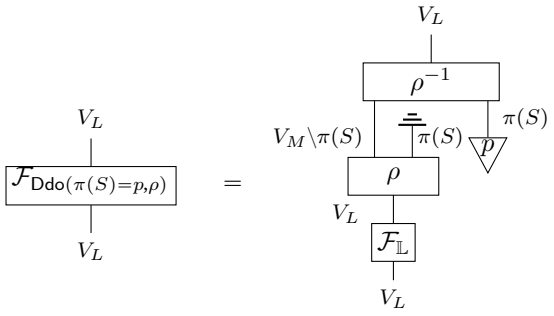
定义 44 （同构构造性/交换抽象）。 所谓同构构造性 （分别为同构交换）抽象 ，是指上述（43）中的设置，其中 (ω, τ) 由构造性抽象（分别为交换抽象）所诱导。

同构构造性抽象。 值得更详细地说明同构构造性抽象的含义。设 (π, τ) 是诱导出 (ω, τ), $S \subseteq V_H$ 的构造性抽象的数据，而 p 是 $\pi(S)$ 的一个尖锐状态。于是，复合向上抽象将 $\mathbb{H}$ 上的 Do-干预 $\mathsf{Do}(S = \tau \circ p)$ ，通过对应的 $\mathbb{M}$ 上的 Do-干预 $\mathsf{Do}(\pi(S) = p)$ ，关联到 $\mathbb{L}$ 上形如 $\rho^{-1}(\mathsf{Do}(\pi(S) = p))$ 的干预；我们将其称为分布式 *Do*-干预，其对任意 $S' \subseteq V_M$ 的一般定义如下：

$$\mathsf{Ddo}(S' = p, \rho) := \rho^{-1}(\mathrm{Do}(S' = p))$$

[30] 他们提出了一种“ 交换干预的分布式版本”，但没有推导其形式，也没有 a
关联的定义 的抽象概念。

因此，我们关注对 $\mathbb{L}$ 施加具有 $S' = \pi(S)$ 的分布式 Do-干预。可以计算得出（证明见附录 A），这些经干预的低层模型的并行机制通道如下。



(45)

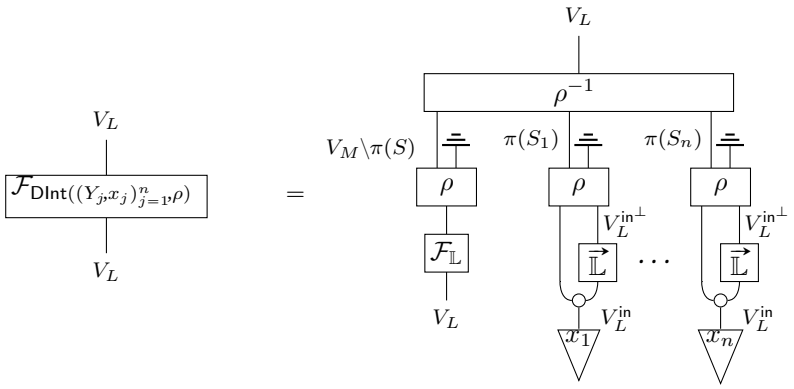
同构构造性抽象产生了一类特征明确的精确变换，通常具有真正的“分布性”。[31] 事实上，根据构造，它属于强类型。

命题 45。 对于同构构造性抽象，整体向上抽象 $(\tau \circ \rho,\ \omega \circ \rho)$ 从 $\mathbb{L}$ 到$\mathbb{H}$ 是强因果抽象（参见定义 *23*）。

同构交换抽象。 类似地，对于同构交换抽象，设 (π, τ) 为诱导出 (τ, ω)的交换抽象的不相交变量对齐；设 $S_1, \ldots, S_n$为 V_H和$\{x_1, \ldots, x_n\}$中互不相交的输入状态，其中 $V_{\text{in}L}$。那么，复合向上抽象将 $\mathbb{H}$上的交换干预 $\mathsf{Inc}((S_j, \tau \circ \rho_{\mathsf{in}} \circ x_j))_{j=1}^n$，经由 $\mathbb{M}$ 上的交换干预 $\mathsf{Inc}((\pi(S_j), \rho_{\mathsf{in}} \circ x_j))_{j=1}^n$，关联到$\mathbb{L}$ 上的形如 $\rho^{-1}(\mathsf{Inc}((\pi(S_j), \rho_{\mathsf{in}} \circ x_j))_{j=1}^n)$ 的干预。我们将其称为分布式交换干预，并对 V_M的任意不相交子集 $Y_1, \ldots, Y_n$作一般性定义：

$$\mathsf{DInt}((Y_j, x_j)_{j=1}^n, \rho) := \rho^{-1}(\mathsf{Inc}(Y_j, \rho_{\mathsf{in}} \circ x_j)_{j=1}^n)$$

因此，我们关注在 $\mathbb{L}$ 上针对 $Y_j = \pi(S_j)$进行的分布式交换干预。令 $S := \cup_j S_j$，再次可以计算出这些受干预低层模型的并行机制通道为：



(46)

我们在附录 A 中证明这一点。

同构抽象的一致性。 最后请注意，上文我们根据并行机制通道展示了低层级“分布式”干预所产生的模型，但没有明确说明整体向上抽象的一致性条件。后者将模型视为从 $\mathbf{V}^{\mathsf{in}}$ 到 $\mathbf{V}^{\mathsf{out}}$的输入输出过程，而前者只能通过 $\mathcal{F}_{\mathbb{M}}$ 通道来明确表示。二者之间的关系需要像引理 40 中那样，对 $\mathcal{F}_{\mathbb{M}}$ 通道的不动点进行量化。

实际上，有一种方法可以将两个视角之间的关系呈现为“结构性的”，从而允许使用图表式一致性条件，例如同构构造性抽象的一致性条件，而无需明确提及不动点。不过，由于这种方法需要稍多一些形式体系，我们将在附录 D 中介绍。

[31]这一　因此，这促使我们提出精确变换的一般概念（定义 21），超越单纯的 Do-干预。　.

无论如何，给定同构构造性抽象的候选数据后，我们可以按如下方式确定它是否成立。对于每个高层干预 $\omega(\sigma) = \mathsf{Do}(S = \tau \circ p)$，我们首先考虑 $\overrightarrow{\mathbb{H}_{\omega(\sigma)}}$ 以及式 (44) 右端的复合。另一方面，通过 $\rho^{-1}(\sigma) = \rho^{-1}(\mathsf{Do}(\pi(S) = p))$，我们可以经由通道 (45) 计算诱导模型 $\mathbb{L}_{\rho^{-1}(\sigma)}$，并重构 $\overrightarrow{\mathbb{L}_{\rho^{-1}(\sigma)}}$ 以及式 (44) 左端的复合，然后在不同输入下将其与右端进行比较。对于同构交换抽象，也可以采用类似的步骤。

6 组件级抽象

到目前为止，我们一直在查询层级考察抽象关系。然而，范畴论视角也启发我们研究更强的抽象概念，它们适用于组合模型的结构 层级，即针对模型的各个组件。本节中，我们将所有组合模型都视为 $\mathbf{C}$ 中的模型，并假定它们具有给定的结构和 查询签名 $(\mathbb{M}, \mathcal{S}, \mathcal{Q}, \mathbf{C})$，其中所有查询都是抽象的（定义 9）。

定义 46 （组件级抽象）。 一个 组件级抽象：

$$(\mathbb{L}, \mathcal{S}_L, \mathcal{Q}_L) \underset{\pi,\, \pi^S}{\overset{\tau}{\rightleftarrows}} (\mathbb{H}, \mathcal{S}_H, \mathcal{Q}_H)$$

由函子 π, π^S 和一个如下所示的满自然变换 τ 给出，其中 π 将 $\mathcal{Q}_H$ 中的查询映射到 $\mathcal{Q}_L$ 中的查询，并且上方的方形可交换：

$$\begin{array}{ccc} \mathbf{Q}_L & \xleftarrow{\ \pi\ } & \mathbf{Q}_H \\ {\scriptstyle |-|_{\mathbb{L}}}\downarrow & & \downarrow{\scriptstyle |-|_{\mathbb{H}}} \\ \mathbf{S}_L & \xleftarrow{\ \pi^S\ } & \mathbf{S}_H \\ & {\scriptstyle [-]_{\mathbb{L}}} \searrow \quad \overset{\tau}{\Longrightarrow} \quad \swarrow {\scriptstyle [-]_{\mathbb{H}}} & \\ & \mathbf{C} & \end{array} \tag{47}$$

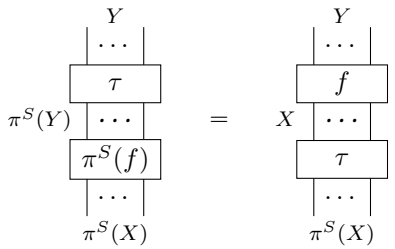
当 τ =id时，组件级抽象是严格 的。

正如术语所示，任何此类数据确实也是查询层级抽象的一个特例，以下是一个直接的结果。

命题 47。 任意组件级抽象 (47) 都会诱导出一个向下抽象 $(\mathbb{L},\ \mathcal{Q}_L) \overset{\tau'}{\rightleftarrows} (\mathbb{H},\ \mathcal{Q}_H)_{\pi}$其中 $\tau' :=\tau \circ 1_{|-|_{\mathbb{H}}}$。

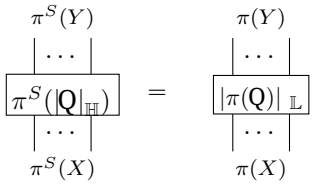
证明。 附录 A。 □

具体而言，组件级抽象会为每个高层变量 $\mathbf{X}$ 指定一个低层变量列表 $\pi^S(\mathbf{X})$，这些变量来自 $\mathcal{S}_L$，并指定一个（确定性的）通道 τ: $\pi^S(X) \to X$ （位于$\mathbf{C}$中），再按通常方式通过乘积将其扩展到高层变量列表（13）、（14）。接下来，它将每个高层组件 $\mathbf{f}$ 映射到低层结构范畴$\mathbf{S}_L$中的一个低层图$\pi^S(\mathbf{f})$，使得以下关系在$\mathbf{C}$中成立：



那么， π^S 会自动扩展为一个函子，而 τ 会自动扩展为一个自然变换。换言之，上述结论对$\mathbf{S}_H$中的任意图表 $\mathbf{f}$ 都成立，也就是说，对分量的复合都成立。

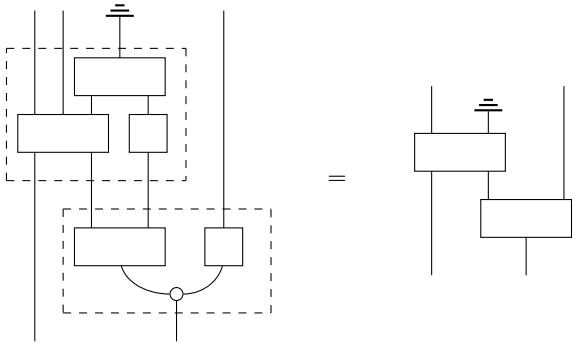
此外，按照惯例，我们还有一个从查询和类型的高层到低层的映射 π ，它将 Q: X $\to$ Y 映为π(Q): π(X) $\to$ π(Y)。最后，分量映射保持这些（抽象）查询对应的图表，即对于每个这样的 Q，我们在低层结构范畴$\mathbf{S}_L$中有以下等式：



总而言之，组件级抽象不仅会将高层查询映射为在 $\mathbf{C}$ 中具有一致表示的低层查询，如今还会附带一种将高层形式图映射为低层形式图的一致映射。此外，将其应用于高层查询的图表时，我们会得到一个与之对应的低层查询所对应的完全相同的低层图表。

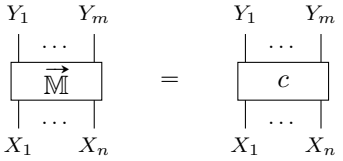
6.1 严格的组件级抽象

一个重要的特殊情形是严格的组件级抽象，其中自然变换是平凡的，且 τ =id。尽管我们最后才定义它们，但它们或许是将高层模型与低层模型联系起来的最简单方式。在这里，我们只需将每个高层组件直接“拆箱”、“细化”或“实现”为一个低层组件图表（同时遵循与查询相关联的图表）。



示例 48。

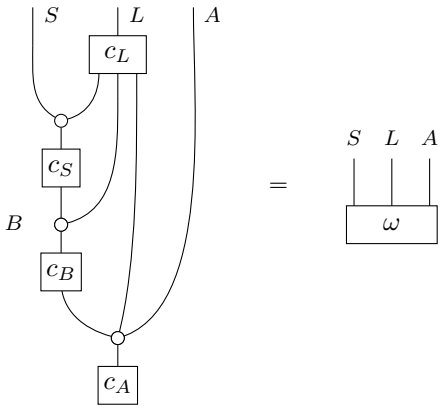
1. 我们将通道 c 在 $\mathbf{C}$ 中的开放因果模型 $\mathbb{M}$ of 定义为具有 $\mathbb{M} = c$ 的模型。



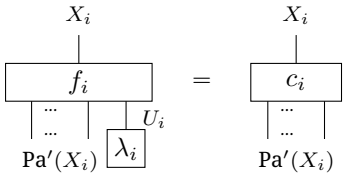
等价地，存在一个严格的组件级抽象 $\mathbb{M} \rightleftarrows$ c ，它保留输入和输出。这里我们假设两个模型都只有一个查询 io: in $\to$ out，分别发送给 $\mathbb{M}$ 和 c ；并且 c 具有变量 X_i,Y_j，以及一个组件 c ，该组件以抽象方式表示 io。[32]

例如，考虑将分布 ω 视为**FStoch**中的一个状态。那么， ω 在$\mathbf{C}$ 中的因果模型，就是其网络图组合后得到 ω 的模型，也就是说，这是通常所说的“针对某个分布的因果模型”。更具体地说，考虑如下分布 ω ，其中 S 表示某人是否吸烟，L 表示其是否患上肺癌， A 表示其年龄。下方左侧的因果模型额外包含一个表示社会经济背景条件的潜变量 B ；当以下等式成立时，它就是 ω 的因果模型（参见 *[CB20*、*LT23]*）。

[32]因此，只有当 $\mathbb{M}$ 和 c各自只有一个输出时， c 本身才是一个因果模型。

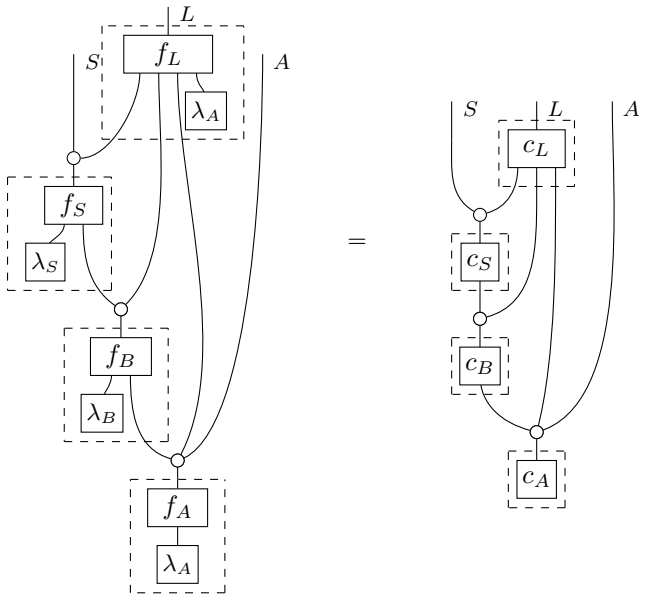


2. 严格的组件级抽象也可以刻画我们如何将 *FCM* 视为对因果模型的"细化"。通常认为，具有内生变量 $E = \{X_i\}_{i=1}^n$ 和外生变量 $U = \{U_i\}_{i=1}^n$ 的 *FCM* $\mathbb{F}$ （定义 *32*）表示支撑变量 E 上的（非函数式）因果模型 $\mathbb{M}$ 的（候选）确定性机制 f_i；其中所有随机性都源于我们对变量 U 的无知，并由相应的"噪声"分布 λ_i 刻画。事实上，给定任意 *FCM* $\mathbb{F}$，我们可以通过为每个 $X_i \in E$ 指定以下定义的机制 c_i，来定义一个仅涵盖 E 的因果模型 $\mathbb{F}|_E$：



(48)

其中$\mathrm{Pa}'(X_i) = \mathrm{Pa}(X_i)\backslash\{U_i\}$ 表示$\mathrm{Pa}(X_i)$ 相对于 $\mathbb{F}$的父集。反过来，给定一个定义在 E 上的（非函数式）因果模型 $\mathbb{M}$，通常有许多种方式可以将其细化为一个*FCM* $\mathbb{F}$，使得 $\mathbb{F}|_E = \mathbb{M}$。由此可得一个严格的组件级抽象 $(\mathbb{F},\ \mathcal{S}_{\mathbb{F}},\ \mathsf{Do(E)}) \rightleftarrows (\mathbb{M},\ \mathcal{S}_{\mathbb{M}},\ \mathsf{Do(E)})$。在此，我们将两个模型中的$\mathsf{X_i}$ 和 X_i 视为相同，并将每个$\mathsf{c_i}$ 映射为$\mathsf{f_i}$ 与 λ_i 的复合（按语法记作），从而使(48) 在$\mathbf{C}$中成立。我们可以如下图示意这种关系，其中同名变量及对应的虚线框均视为相同。

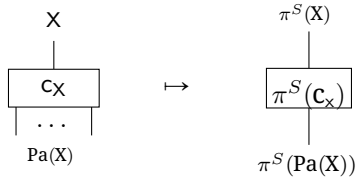


(49)

6.2 机制级因果抽象

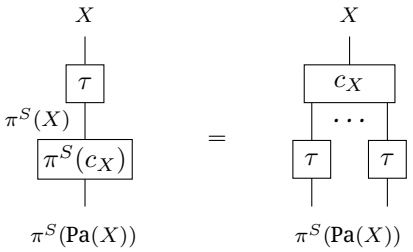
原则上，我们可以探讨第5节中的任意一种因果向下抽象何时能进一步扩展到组件级。本节将针对构造性抽象进行探讨，并将能够扩展为组件级抽象的构造性因果抽象称为机制级因果抽象。由此，我们得到了一种新颖且更强的因果抽象概念，这一概念此前尚未在文献中得到探讨。[33]

具体而言，考虑一个构造性抽象，即向下抽象 $(\mathbb{L},\ \mathbf{Do}(\mathbf{V}_L)) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H},\ \mathbf{Do}(\mathbf{V}_H))$（定理25）。要形成机制层面的因果抽象，就意味着存在一个进一步的映射π^S，将每个高层机制映射到一个低层图：



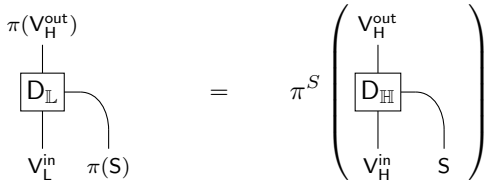
(50)

suc因此以下条件必须在$\mathbf{C}$ 中成立（按照惯例，使用我们的字体约定省略 $[-]_{\mathbb{L}/\mathbb{H}}$ ）：



(51)

接着，我们将 π^S 自由扩展为一个从高层图表到低层图表的函子，使得对所有 $X \in \mathbf{V}_H$ 都有 $\pi^S(X) = \pi(X)$。此外，对于每个子集 $\mathbf{S} \subseteq \mathbf{V}^{\mathsf{in}^\perp}_{\mathsf{H}}$，我们都有一个$\mathbf{S}_L$中的等式，且该等式纯粹存在于图表层面：



(52)

也就是说，在 $\pi(S)$y π^S 处对低层模型执行开放操作，得到的形式图表，与将 π^S 应用于高层模型在 S 处执行开放操作的图表所得结果相同。

我们很快就会看到，在对模型非常温和的假设下，映射 π^S 只要存在就是唯一的，因此我们可以将其视为抽象 $(\pi、\ \tau)$的一个性质，正如我们的术语所暗示的那样。为此，下面这个纯粹的图论定义将会有用。

定义49. 设 π 是构造性因果抽象 $(\pi,\ \tau)$ 在 $\mathbb{L}$ 与 $\mathbb{H}$之间的划分，并对每个$\mathsf{X} \in \mathbf{V}_H$ 定义：

$$\alpha(\mathsf{X}) := \{\mathsf{Z} \in \mathbf{V}_L \mid \exists\ \textit{a directed path } Z \to \pi(\mathsf{X})\ \textit{in } G_L\ \textit{which does not pass through } \pi(\mathrm{Pa}(\mathsf{X}))\}$$

我们称 π 为：

- simple *if* $\alpha(\mathsf{X}) \cap \pi(\mathsf{Y}) = \varnothing$ *for* $\mathsf{X} \neq \mathsf{Y}$*;*
- extra-simple[35] *if* $\alpha(\mathsf{X}) \cap \alpha(\mathsf{Y}) = \varnothing$ *for* $\mathsf{X} \neq \mathsf{Y}$*;*

[33]这也推广了例48（2），它是 $\tau = \mathsf{id}$的一个实例。[34]这里我们利用了 $\mathcal{S}_L$ 的变量是$\mathbf{Do}(\mathbf{V}_L)$的类型这一事实。注意，如前所述，自然性条件（51）随后会自动推广到任意高层图表。[35]注意，根据定义， $\pi(\mathbf{X}) \subseteq \alpha(X)$ 。

• 满的 当对于所有 $\mathsf{Y} \in \pi(\mathrm{Pa}(X)\backslash \mathsf{V}_{\mathsf{inH}})$，都存在一条有向路径 $\mathsf{Y} \to \pi(X)$ 通向 G_L。

直观而言，$\alpha(X)$ 刻画了 $\mathbf{Pa}(X)$ 的低层表示在多大程度上"屏蔽"了 X 的低层表示。现在，我们可以为 $(\pi$、$\tau)$ 的条件命名，以得到图表（52）中的等式。事实上，我们将把若干结果合并呈现，因为这些条件将取决于如何为每个模型定义由其签名 $\mathcal{S}$ 生成的结构范畴 $\mathbf{S}$；对此我们将考虑几种选择，并分别给出相应的结果。

对于这些选项中的第一个，我们考虑对例如 [CJ19] 中马尔可夫范畴概念的轻微弱化。*cd*-范畴 $\mathbf{C}$ 的定义与马尔可夫范畴相同，但每个对象 A 上选定一个（而非唯一的）丢弃态射 $_A\colon A \to I$，并满足：

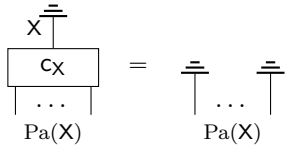


cd-函子 $F\colon \mathbf{C} \to \mathbf{D}$ 还必须保持这些丢弃态射。[36]

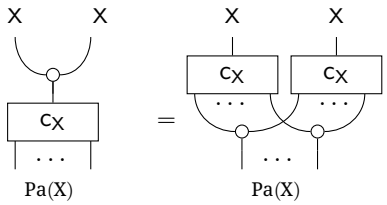
定义 50。给定因果模型 $\mathbb{L}, \mathbb{H}$，其对应的自由结构范畴分别为 $\mathbf{S}_L, \mathbf{S}_H$，若一组 $(\pi(\mathsf{X}))_{X \in \mathsf{V}_\mathsf{H}}$ 由 V_L 的不相交子集构成，则称其为结构良好，如果 π 满足：

• 在由 $\mathcal{S}$ 生成的自由 *cd* 范畴 $\mathbf{S}$ $=\mathsf{CD}(\mathcal{S})$ 中是额外简单且满的（参见 *[LT23、TLC+24]*，本质上也参见 *[JKZ19]*）；

• 在自由马尔可夫范畴 $\mathbf{S} = \mathsf{Markov}(\mathcal{S})$ 中是额外简单的；等价地，我们添加公理，规定每个机制都是一个通道（参见 *[FL23]*）：



• 简单的，在使用自由笛卡尔范畴 $\mathbf{S}$ $= \mathsf{Cart}(\mathcal{S})$ 时，我们等价地添加更多公理，说明每个机制都是确定性信道：



若每个 $\pi^S\colon \mathbf{S}_H \to \mathbf{S}_L$ 结构良好，我们称 *cd*-函子 $\pi^S(\mathsf{cX})$ 为结构良好的，当且仅当它是 $\mathbf{S}_L$ 中的规范化网络图；并且在每种情况下，若 $(\pi(\mathsf{X}))_{X \in \mathsf{V}_\mathsf{H}} := (\pi^S(\mathsf{X}))_{X \in \mathsf{V}_\mathsf{H}}$ 是不相交子集，且 $\pi(\mathsf{V}_\mathsf{H}^{\mathsf{in}}) = \mathsf{V}_\mathsf{L}^{\mathsf{in}}$、$\pi(\mathsf{V}_\mathsf{H}^{\mathsf{out}}) \subseteq \mathsf{V}_\mathsf{L}^{\mathsf{out}}$ 结构良好。

沿着定义 50 的列表往下，我们在$\mathbf{S}$中加入更多图式公理，因此等式（52）更容易成立。现在我们能够证明刻画机制级因果抽象的主要结果。

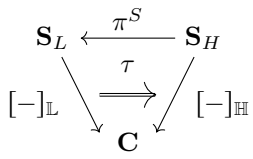
定理 51. 设 $\mathbb{L}, \mathbb{H}$ 是 $\mathbf{C}$ 中的因果模型，且满足 $\mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{H} \subseteq \mathsf{V}_\mathsf{H}^{\mathsf{out}}$。假设$\mathbb{L}$ 的$\mathsf{V}_\mathsf{L}^{\mathsf{in}}$中每个输入变量在$\mathbf{C}$中至少有一个规范化状态。对于每种结构类型（笛卡尔、马尔可夫或 *cd*），以下条件等价：[37]

1. 一种机制级因果抽象 $(\pi,\ \pi^S,\ \tau)$；

[36]在 cd-范畴中，我们将态射 f 称为 *channel*，当且仅当$f = _\circ f$。cd-函子 F 和自然变换 α 与马尔可夫情形类似，但还要求 $F(_A) = _{F(A)}$ 以及所有变换分量 α_X（对于 F，还包括结构同构）都是通道。[37]更一般地，对于该结果，我们只要求：对于每个输入变量 $\mathsf{V} \in \mathsf{V}^{\mathsf{in}}$ 和所有 f、g，在$\mathbf{C}$中都有 $X \to Y\ _V \otimes f = _V \otimes g \implies f = g$。

2. 一种构造性抽象 $(\pi,\ \tau)$，使得 π 结构良好；

3. 一个结构良好的 *cd*-函子 π^S 及如下自然变换 τ：



证明。附录A.1。 □

该结果表明，当划分 π 满足相应条件时，构造性抽象当且仅当其划分满足相应条件，便可扩展到机制层面：根据结构类型的选择，条件可以是简单、额外简单，或额外简单且满。虽然我们针对 $\mathbf{S}$ =CD($\mathcal{S}$)给出了一个结果，但对于典型的因果模型，使用自由马尔可夫范畴 $\mathbf{S}$ =Markov($\mathcal{S}$)或许更自然，因为我们知道每个机制都是 $\mathbf{C}$中的一个通道，而在这种情况下，π 必须是额外简单的。不过，对于确定性模型（例如神经网络），采用 $\mathbf{S}$ =Cart($\mathcal{S}$)更为自然，此时我们只要求 π 是简单的。

等价地，最后一个条件仅用结构层级上的函子来表达这种抽象。使用自由马尔可夫范畴作为结构时，这个最后条件可以用一种更简单的形式来表述。

引理 52。令 $\mathbb{L}$、$\mathbb{H}$ 为满足 $\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}} \subseteq \mathsf{V}^{\mathsf{out}}_{\mathsf{H}}$ 的因果模型。*cd*-函子 π^S：马尔可夫($\mathcal{S}_{\mathbb{H}}$) → 马尔可夫($\mathcal{S}_{\mathbb{L}}$) 具有结构良好性，当且仅当 $\pi^S(\mathsf{D}_{\mathbb{H}})$ 在 $\mathbf{S}_L$ 中再次成为网络图，且满足 $\pi^S(\mathsf{V}^{\mathsf{in}}_{\mathsf{H}}) = \mathsf{V}^{\mathsf{in}}_{\mathsf{L}}$、$\pi^S(\mathsf{V}^{\mathsf{out}}_{\mathsf{H}}) \subseteq \mathsf{V}^{\mathsf{out}}_{\mathsf{L}}$以及 $\pi^S(\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}}) \subseteq \mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{L}}$。

证明。 附录A.1。 □

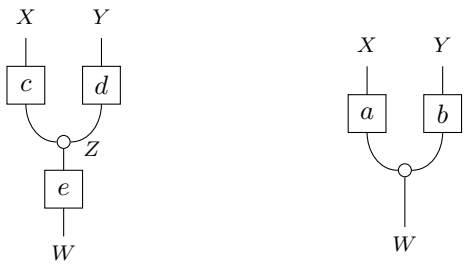
因此，在使用自由马尔可夫范畴时，指定一个机制级因果抽象等价于给出一个 cd-函子 π^S 和变换 τ，它们按照上述意义保持整体网络图。或许令人意外的是，这告诉我们，可以完全根据机制和输入输出行为来定义这一潜在的广泛构造性抽象类别，从而完全不必明确涉及干预或（Do-）查询。于是，一个开放问题是：实践中所有有趣的构造性抽象案例是否都属于这种形式，即其划分 π 是简单的。

现在我们给出机制级因果抽象的一些例子与反例。

例 53. 构造性抽象的例 *27* 在取 $\mathbf{S} = \mathsf{Cart}(\mathcal{S})$ 时实际上是机制层面的；由于该模型是确定性的，这是一个恰当的选择。事实上，在取$\mathbf{S}$ = 马尔可夫($\mathcal{S}$) 时，它更强地也是如此。借助定理 *51* 的 (1) ⇔ (2) 很容易验证这一点，因为 π 显然是简单的，而且事实上也不难看出它还是额外简单的。[38]

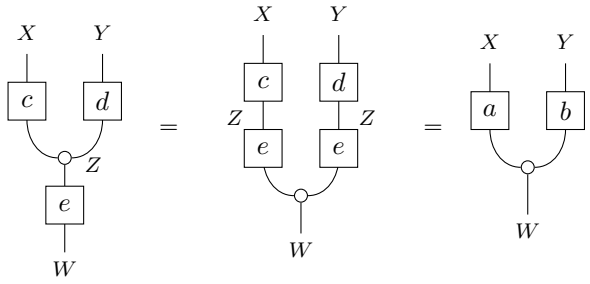
下面给出更多教学示例，用以说明定义 50 的不同情形。

例 54. 设 $\mathbb{L}$ 与 $\mathbb{H}$ 为 $\mathbf{C}$ 中由下列各自的网络图所定义的因果模型：

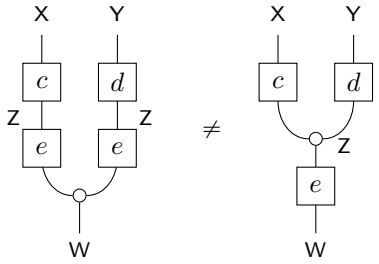


[38]注意，例 27 中式 (24) 的条件使得我们也可以借助定理 51 的 (1) ⇔ (3) 以及引理 52，更直接地在机制层面建立该关系。$\pi^S(\mathsf{D}_{\mathbb{H}})$ 是一张网络图这一点是显然的。

假设在 $\mathbf{C}$ 中，态射 a, b, c, d, e 均为确定性的，且 $a = c \circ e$ 与 $b = d \circ e$。则存在一个构造性抽象，其 $\tau = \mathrm{id}$ 由如下设定给出：$\pi :: \mathsf{X} \mapsto \{\mathsf{X}\}, \mathsf{Y} \mapsto \{\mathsf{Y}\}, \mathsf{W} \mapsto \{\mathsf{W}\}$。例如，对于 $\overrightarrow{\quad}$ $\overrightarrow{\quad}$，我们得到高层输入输出通道 $\mathbb{H} = \mathbb{L}$，因为：



这个抽象并非额外简单，因为对于此 π，我们有 $\alpha(\mathsf{X}) = \{\mathsf{X}, \mathsf{Z}\}$、$\alpha(\mathsf{Y}) = \{\mathsf{Y}, \mathsf{Z}\}$ 以及 $\alpha(\mathsf{W}) = \{\mathsf{W}\}$，因此 $\alpha(\mathsf{X})$ 和 $\alpha(\mathsf{Y})$ 并不互不相交。若在结构中使用自由马尔可夫范畴 $\mathbf{S}$ =马尔可夫($\mathcal{S}$)（或自由 $\boldsymbol{cd}$ 范畴），便会发现无法实现组件级抽象，因为可以得出 $\pi^S(\mathsf{D}_{\mathbb{H}}^{X\,Y}) \neq \mathsf{D}_{\mathbb{L}}^{X\,Y}$，。事实上，将其视为图表，我们有：

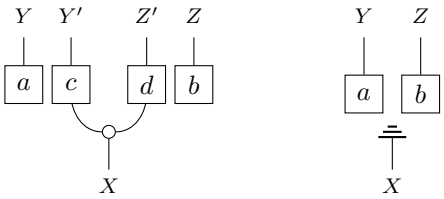


(53)

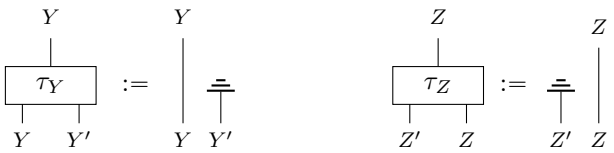
尽管如此，这个抽象是简单的；如果结构改用自由笛卡尔范畴$\mathbf{S}$ = 笛卡尔($\mathcal{S}$)，也就是要求每个盒子都是确定性的，那么(53) 就变成一个等式，我们也就能得到组件级抽象。

请注意，也可以用另一种方式定义构造性抽象，同样使用 $\tau = \mathrm{id}$，令 $\pi :: \mathsf{X} \mapsto \{\mathsf{X}\}$、$\mathsf{Y} \mapsto \{\mathsf{Y}\}$、$\mathsf{W} \mapsto \{\mathsf{W}, \mathsf{Z}\}$。现在，对于每个$\mathsf{N}$ =X、Y、Z，都有 $\alpha(\mathsf{N}) = \pi(\mathsf{N})$，从而使该抽象成为额外简单的抽象；即使使用自由马尔可夫范畴，这也定义了一种机制级因果抽象。

示例 55。令 t $\mathbb{L}$ 和 $\mathbb{H}$ 是因果模型，分别由 $\mathbf{C}$ 中的以下网络图定义：



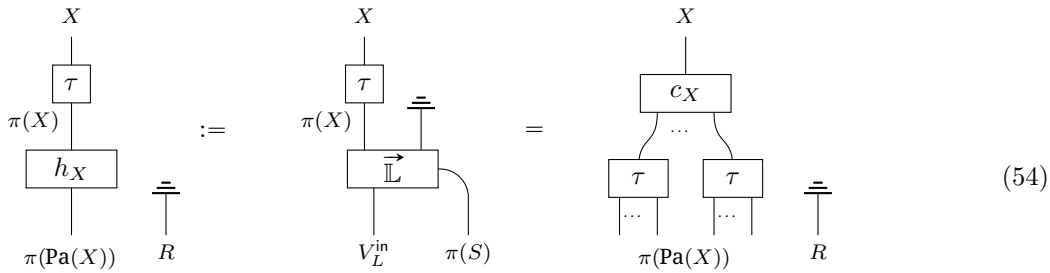
$\mathbb{L}$ 在给定 $\pi :: \mathsf{X} \mapsto \{\mathsf{X}\}$，$\mathsf{Y} \mapsto \{\mathsf{Y}, \mathsf{Y}'\}$，$\mathsf{Z} \mapsto \{\mathsf{Z}, \mathsf{Z}'\}$ 和（$\tau_X = id_X$）的情况下，$\mathbb{L}$ 与 $\mathbb{H}$ 之间存在一个构造性抽象：



然而，这并不简单，因为 $\mathsf{X} \in \pi(\mathsf{X}) \cap \alpha(\mathsf{Y})$。[39] 这种抽象无法扩展为组件级抽象；事实上，我们甚至无法像 (50) 那样定义 π^S，因为 $\pi^S(\mathsf{a})$ 会是$\mathsf{Y} \otimes \mathsf{Y}'$ 在$\mathbf{S}_L$中的一个状态，但不存在这样的状态。

[39]这意味着条件（63）不成立，因此 π^S 确实不可定义。

备注 56。 请注意，任何构造性抽象确实都意味着每个高层机制都满足一个接近公式(51)的条件。对于任意$\mathsf{X} \in \mathsf{V}^{\mathsf{in}^\perp}_{\mathsf{H}}$，令$\mathsf{S} := \mathrm{Pa}(X)\backslash\mathsf{V}^{\mathsf{in}}_{\mathsf{H}}$。那么在$\mathbf{C}$ 中，我们有下面的右侧等式：



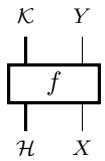
(54)

其中 $R = \mathsf{V}^{\mathsf{in}}_{\mathsf{L}}\backslash\pi(\mathrm{Pa}(X))$。接下来，我们希望定义一个规范化网络图 h_{X}，如左端所示，作为 $\pi^S(\mathsf{c}_{\mathsf{X}})$ 的候选。要使这样的规范化图存在，并以 $\pi(\mathrm{Pa}(X))$ 作为输入（这是π^S 构成函子所必需的），就意味着从 $\mathsf{V}^{\mathsf{in}}_{\mathsf{L}}$ 中某个输入到 $\pi(X)$ 的每条路径都经过 $\pi(\mathrm{Pa}(X))$，也就是说， $\alpha(X) \cap \mathsf{V}^{\mathsf{in}}_{\mathsf{L}} = \varnothing$。因此，这与一开始定义函子 π^S 的条件等价。要证明在每种情况下 π^S 都满足机制级因果抽象的条件，恰恰需要定义 ***50*** 中更强的条件；详情请参见附录***A.1***中的证明。

7 量子模型的抽象

我们对抽象所作的范畴论定义有一项优势：它不仅适用于因果模型，也适用于任何组合模型。另一大类模型是基于量子过程的模型。本节将介绍量子模型与经典模型之间的新型抽象概念，并探讨其在可解释量子人工智能中的潜在应用。为此，我们将以下面的终端$\mathbf{d}$-范畴为例展开讨论；它既能刻画有限维量子过程（如量子计算），也能刻画有限经典概率过程（尤其是经典计算） [CK18]。

例 57 (**量子-经典**)。 在范畴 **量子-经典** 中，对象是成对的 $(\mathcal{H}, X)$，其中 $\mathcal{H}$ 是有限维复希尔伯特空间（量子系统），X 是有限集（经典系统）；它们分别用粗线和细线表示，并排放置。态射 $f\colon (\mathcal{H}, X) \to (\mathcal{K}, Y)$，可绘制为：



由一个受控量子仪器给出，即一组线性映射 $(f(y \mid x)\colon L(\mathcal{H}) \to L(\mathcal{K}))_{x\in X, y\in Y}$，其中每个映射都是完全正 的，且联合保迹：

$$\sum_{y\in Y} \mathrm{Tr}(f(y \mid x)(a)) = \mathrm{Tr}(a)$$

对于所有 $x \in X$ 和算子 $a \in L(\mathcal{H})$。此处， $L(\mathcal{H})$ 是 $\mathcal{H}$上所有线性算子的集合。[40] 给定任意$g\colon (\mathcal{K}, Y) \to (\mathcal{W}, Z)$ 我们通过求和定义 $g \circ f\colon (\mathcal{H}, X) \to (\mathcal{W}, Z)$：

$$(g \circ f)(z \mid x) := \sum_{y\in Y} g(z \mid y) \circ f(y \mid x)$$

我们定义 $(\mathcal{H}, X) \otimes (\mathcal{K}, Z) := (\mathcal{H} \otimes \mathcal{K}, X \otimes Y)$，其中 $\mathcal{H} \otimes \mathcal{K}$ 是希尔伯特空间的张量积；在态射上定义 $(f \otimes g)(y, z \mid x, w) := f(y \mid x) \otimes g(z \mid w)$，单位对象为 $I = (\mathbb{C}, \{\star\})$。此范畴具有丢弃映射：$(\mathcal{H}, X) \to (\mathbb{C}, \{\star\})$，其中 $(\star \mid x)$：$L(\mathcal{H}) \to \mathbb{C}$定义为对所有 $a \in L(\mathcal{H})$均有 $a \overset{\star}{\mapsto} \mathrm{Tr}(a)$。根据定义，每个态射都是通道，因为它满足迹保持性。

[40]回顾一下，线性映射 $f\colon L(\mathcal{H}) \to L(\mathcal{K})$ 若将正算子 $a = b^\dagger b \in L(\mathcal{H})$ 映射为 $L(\mathcal{K})$中的正算子 $f(a)$，则称其为正映射；若对于任意其他希尔伯特空间 $\mathcal{H}'$，映射 $f \otimes \mathrm{id}_{\mathcal{H}'}$ 也为正映射，则称其为完全正映射。

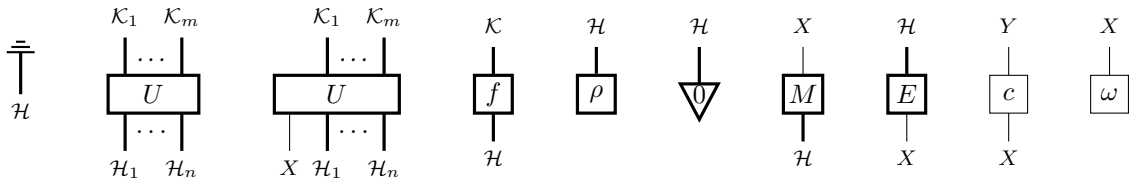


图 2：**量子-经典**中的态射，依次为： $\mathcal{H}$上的偏迹；酉门，即幺正算子U的CPTP映射 $U(-)U^\dagger$，作用于多个量子系统；带经典控制 X的受控酉算子 U；CPTP映射 f： $L(\mathcal{H}) \to L(\mathcal{K})$；密度矩阵 $\rho \in L(\mathcal{H})$；计算基中的纯态 $|0\rangle\langle 0|$；在 $\mathcal{H}$ 上、具有有限结果集 X的测量 M；从有限集 X 到 $\mathcal{H}$的编码器 E；概率通道 c： $X \to Y$； X上的概率分布 ω。

在此范畴中，“全量子”系统 $\mathcal{H} = (\mathcal{H}, \{\star\})$对应有限维希尔伯特空间（粗线），例如量子比特 $\mathcal{H} = \mathbb{C}^2$；“全经典”系统 $X = (\mathbb{C}, X)$则对应有限集（细线）。量子信道，即量子系统之间的信道 f: $\mathcal{H} \to \mathcal{K}$，恰好是完全正且保迹（*CPTP*）的映射，其中对应于偏迹。这些包括量子态和量子酉电路。有限集之间的经典信道 c: $X \to Y$ 恰好是 **FStoch** 中的信道，由此得到嵌入 $\mathbf{FStoch} \hookrightarrow \mathbf{QC}$，使我们能够在 **QC** 中描述经典有限因果模型。图 *2* 展示了 **QC** 中两类系统之间的一些重要过程。

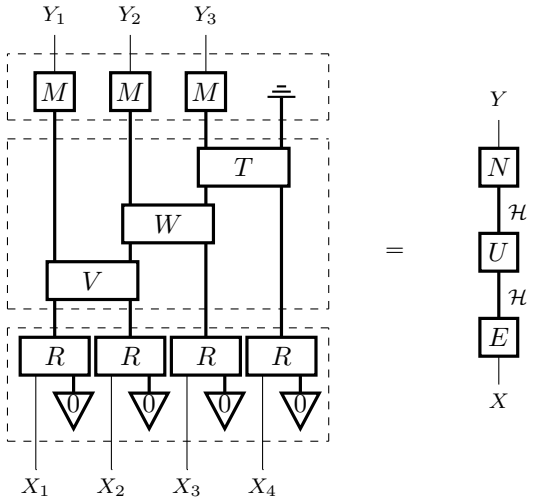
所谓量子模型，是指语义取自**量子-经典**、至少使用某些量子系统 $\mathcal{H}$、并且（可能）也使用经典系统的组合模型 $\mathbb{M}$。所谓经典模型，是指在**量子-经典**中仅使用经典系统的模型 $\mathbb{M}$，例如任何（有限）因果模型。现在，探讨两侧模型各自可为任一种类的抽象关系便有了意义。

7.1 量子到量子抽象

我们首先考虑 $\mathbb{L} \to \mathbb{H}$ 其中两者 $\mathbb{L}$ 和 $\mathbb{H}$ 都是量子模型的抽象关系。一个普遍的例子是严格的组件级抽象，它刻画了量子模型 $\mathbb{H}$ （例如，处于计算层级或逻辑层级）如何通过由特定量子门构成的较低层级量子模型 $\mathbb{L}$（例如，处于量子硬件层级）来实现，如下例所示。尽管从某种意义上说，这个例子很简单，但它表明抽象并非量子理论和量子计算中的陌生概念；我们将对其他更有趣的量子到量子抽象案例的研究留待未来工作。

示例 58 （量子模型实现）。 所谓从经典输入 X到经典输出 Y 的量子电路模型 $\mathbb{M}$，是指**量子-经典** 中的一个模型，它由从 X 到量子系统 $\mathcal{H}$ 的编码器 E、随后作用于 $\mathcal{H}$的酉电路 U，以及带输出Y的最终测量 N 顺序复合而成。在实践中， $\mathcal{H}$ 是若干量子比特的张量，而 $\mathbb{M}$ 由作用于这些量子比特的若干底层过程实现，并由严格的组件级抽象 $\mathbb{L} \rightleftarrows \mathbb{M}$ 刻画，其底层模型$\mathbb{L}$ 包含各个量子比特。在一个示例中， $\mathbb{L}$ 作用于四个量子比特，且 X 分解为四个输入和

Y 作为三个输出如下所示，使用图 *2* 中定义的过程。



编码器 E 通过作用于初始态 $|0\rangle$的受控旋转门 R 实现；幺正算子 U 通过（通常为）纠缠门 V，W，T 实现；测量 N则通过对其中三个量子比特分别进行测量来实现。

7.2 量子到经典抽象

一种值得探索的新可能性是从《Quantum》模型 $\mathbb{L}$ 抽象到高层级经典模型 $\mathbb{H}$的抽象 $\mathbb{L} \to \mathbb{H}$ 。考虑这类抽象有几种动机：

1. 经典约化。一个简单的动机是，精确解释量子模型 $\mathbb{L}$ 如何“本质上”是经典的，即它可以由经典模型 $\mathbb{H}$ 完全说明。当对 $\mathbb{L}$ 的查询足够“丰富”，或者抽象甚至在组件级成立、涵盖我们打算使用量子模型的所有方式时，这一点会最为明显。进行这种约化的一个理由，是反驳这样一种主张：要实现模型的某些行为（例如计算行为，或量子认知 [TO26]等场景中的行为），量子语义是必不可少的；相关实例见附录 E 中的示例 77。

2. 因果退相干。 量子基础领域的另一个相关动机，是描述一种退相干形式：从底层的量子描述中涌现出对物理系统的“因果一致”经典描述。关键在于，这要求 $\mathbb{L}$ 是一个量子因果模型，而量子因果模型并不能直接定义为组合模型；我们将在第 8 节进一步讨论这一未来方向。

3. 解释。 人工智能与量子计算交叉领域的最后一个相关动机，是探讨经典因果模型 $\mathbb{H}$ 如何为量子模型 $\mathbb{L}$ 的（部分）行为提供“解释”。事实上，可解释人工智能（XAI）领域的研究者认为，对低层级神经网络 $\mathbb{L}$ 的理想 “ 解释 ”，最好描述为向高层级（可解释）因果模型 $\mathbb{H}$ [GIZ+23, GRP20, GWP+24] 的因果抽象 $\mathbb{L}$ 。这种抽象以结构方式捕捉 $\mathbb{H}$ 如何反映网络的行为；也就是说，后者必须产生符合 $\mathbb{H}$ 因果结构的结果，从而提供因果解释，而因果解释常被认为是可能存在的“最丰富”的解释形式（另见第 1 节）。

同样地，我们或许可以希望通过将参数化量子电路（PQC）或更一般的量子模型 $\mathbb{L}$ 抽象为经典因果模型 $\mathbb{H}$，来解释它们。这或可为可解释量子人工智能（XQAI）提供一种途径：通过此类抽象，将由参数化量子电路给出的人工智能模型解释为可解释的经典因果模型。请注意，这与完全经典的可解释人工智能情形有所不同：尽管任何 神经网络在形式上都可以被视为因果模型，但量子模型并非如此（参见第 8 节的讨论）。尽管如此，这未必会构成障碍——关于解释（即解释项）的关键在于

解释（即解释项）有两个方面：被假定 为因果模型的高层模型 $\mathbb{H}$，以及这两个模型之间的联系。

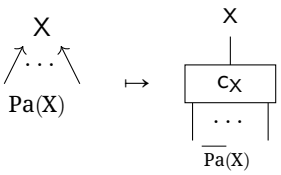
本节剩余部分将聚焦于上述第三个也是最后一个动机。无论如何，抽象的含义和丰富程度，恰恰取决于与高层级查询相关联的低层查询集合。因此，我们必须提出这样的问题：以哪些有意义的方式查询量子电路，而且如何使其能够为经典模型（理想情况下是因果模型）提供有意义的抽象？在此，我们将通过把因果模型推广到一类更广泛、允许采用量子语义的组合模型，并适当地推广我们先前提出的两类查询——抽象 Do 查询和交换查询，来对这个问题给出部分答案。

重新审视有向无环图的模型。为了定义这类组合模型，我们首先给出一种更通用的新方法，用于根据有向无环图定义字符串图。令 G 为一个开放有向无环图，其顶点集合为 V，并令 $\mathrm{V}^{\mathrm{out}} \subseteq \mathrm{V}$ 为另一个顶点子集，称为输出。

我们将图表 $\mathrm{D}(G, \mathrm{V}^{\mathrm{out}})$ 定义为：对于每个顶点 $\mathrm{X} \in \mathrm{V}$，设置一条导线 X；对于 G中的每条边 $\mathrm{X} \to \mathrm{Y}$，设置一条导线 $\mathrm{e}_{\mathrm{X},\,\mathrm{Y}}$；并为每个输出 X 另设一条标记为 $\mathrm{X}^{\mathrm{out}}$ 的导线。接着，对于每个非输入 X，我们绘制由盒子 $s_X \circ c_X$ 组成的复合；对于每个输入 $\mathrm{I} \in \mathrm{V}^{\mathrm{in}}$，则只绘制一个盒子 s_I，如下所示。最后，将这些盒子沿对应的导线连接起来，便得到 $\mathrm{D}(G, \mathrm{V}^{\mathrm{out}})$。



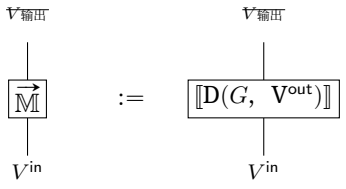
这里 $\overline{\mathrm{Pa}}(\mathrm{X}) := (\mathrm{e}_{\mathrm{Y},\,\mathrm{X}})_{\mathrm{Y} \in \mathrm{Pa}(\mathrm{X})}$ 以及 $\overline{\mathrm{Ch}}(\mathrm{X}) := (\mathrm{e}_{\mathrm{X},\,\mathrm{Y}})_{\mathrm{Y} \in \mathrm{Ch}(\mathrm{X})}$；或者，如果 $\mathrm{X} \in \mathrm{V}^{\overline{\mathrm{out}}}$，那么 $\mathrm{Ch}(\mathrm{X})$ 还额外包含 $\mathrm{X}^{\mathrm{out}}$。因此，$\mathrm{D}(G, \mathrm{V}^{\mathrm{out}})$ 是一个从 V^{in} 到输出 $V_{\mathrm{out}} = (\mathrm{X}^{\mathrm{out}})_{\mathrm{X} \in \overline{\mathrm{V}^{\mathrm{out}}}}$的图表。我们约定，当输出 $\mathrm{X} \in \mathrm{V}^{\mathrm{out}}$ 在 G 中没有子节点时，省略绘制 s_{X}，并将 $\mathrm{X} = \mathrm{X}^{\mathrm{out}}$标识为：[41]



稍后我们将在例63和例64中看到 $\mathrm{D}(G、\mathrm{V}^{\mathrm{out}})$ 的例子。现在可以给出有向无环图模型的一般概念。

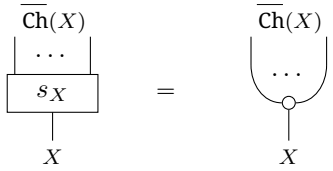
定义 59. 给定一个开放有向无环图 G 及输出顶点的子集 $\mathrm{V}^{\mathrm{out}}$，我们称 带输出的 G组合模型$\mathrm{V}^{\mathrm{out}}$ 在 $\mathbf{C}$ 中，是指一个组合模型 $\mathbb{M}$，其签名为 $\mathcal{S}_{\mathrm{D}(G,\ \mathrm{V}^{\mathrm{out}})}$ 由 $\mathrm{D}(G、\mathrm{V}^{\mathrm{out}})$ 的线标签和盒子给出，并且还将每个盒子映射到 $\mathbf{C}$ 中的一个通道。

为简洁起见，对于每个非输入的 X，我们记 $m_X := s_X \circ c_X$ 属于 $\mathbf{C}$。和之前一样，我们将整体通道记为 $\mathbb{M}$ 从输入到输出 V^{in} 的通道为：



[41]形式上，我们假设一个具有 $X = X^{\mathrm{out}}$ 和 $s_X = \mathrm{id}$ 的模型。

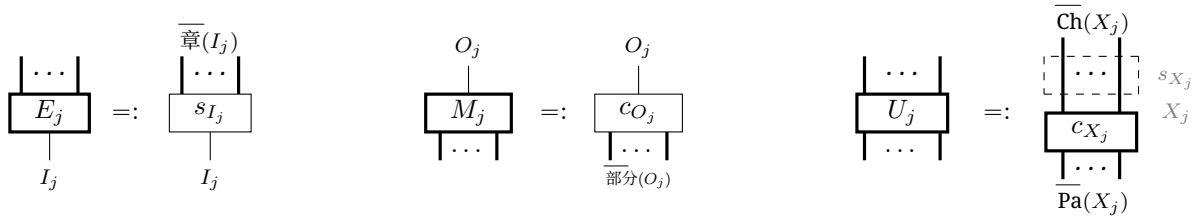
例 60 (经典因果模型**)**。 在马尔可夫范畴 $\mathbf{C}$ 中，具有输出V^{out} （见定义 *10*）的 G 的一个（开放）因果模型，等价于上述意义下的 G 的一个组合模型，并且满足 $e_{X,Y} = X$ 对所有$\mathsf{X},\mathsf{Y} \in \mathsf{V}_{\mathsf{G}}$ 且$\mathsf{X} \in \mathrm{Pa}(\mathsf{Y})$时成立， $X^{\text{out}} = X$ 对所有$\mathsf{X} \in \mathsf{V}^{\text{out}}$ 均成立，并且满足以下条件：



构造 直接对 $\mathsf{D}(G,\ \mathsf{V}^{\text{out}})$ 作相应替换，便得到网络图 D_G。 [42]

示例 61 （量子电路）。 考虑一个量子电路模型 $\mathbb{Q}$ ，其含义如示例 *58* 所述，具有经典输入 X 和经典输出 Y ；此外，该模型已经分解为多个盒子的复合，其中每个盒子要么是编码器 E （即从经典输入到量子），要么是幺正算子 U （从量子到量子），要么是测量 M （从量子到经典输出）。

我们可以将 $\mathbb{Q}$ 视为一个开放有向无环图 $G_{\mathbb{Q}}$ 的组合模型，定义如下。对于每个编码器 E_j的输入，指定一个输入顶点 $\mathsf{I_j}$ ；对于每个测量 M_j 的输出，指定一个输出顶点 $\mathsf{O_j}$ ；对于每个幺正算子 U_j，指定一个顶点 $\mathsf{X_j}$ ，其中 $\mathrm{Pa}(\mathsf{X_j})$ 和 $\mathrm{Ch}(\mathsf{X_j})$ 分别由 U_j的输入和输出给出。因此，该图中的每条边对应于电路的一条内部线路。于是， $\mathbb{Q}$ 是 $G_{\mathbb{Q}}$的组合模型，其输出为 O，通过 $X_j : = \bigotimes_{Y \in \mathsf{Ch}(X_j)} Y$ 、 $s_{X_j} = \mathrm{id}$ 和 $U_j = c_{X_j} = m_{X_j}$给出。 [43]



例 62 （量子因果模型的子类）。 在 G 具有特定意义的情况下，任何具有有向无环图 G 的量子因果模型（假设每个“实验室”满足某种自然的维数条件）都等价于上述意义下的一个关于 G 的组合模型的数据，其对应方式类似于例 *61* 中对幺正门的处理。这些模型是 *[BLO19, ABH+17]* 中更一般的量子因果模型概念的特例；一般而言，这些模型（目前）尚未被定义为组合模型；参见第 *8* 节的讨论。

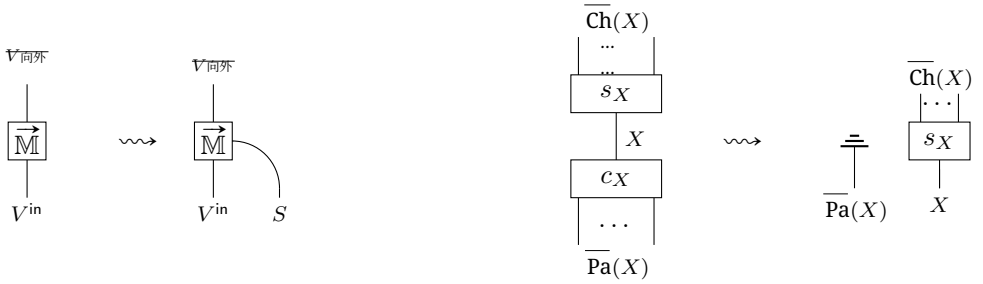
请注意，虽然可以轻易地将任何经典因果模型转变为一个输出子集任意指定的新模型（沿用相同的机制，只需在原网络图中为任何新增输出添加复制映射），但对于定义 59 中开放有向无环图的组合模型，情况就不再如此，因为输出的选择并非无关紧要。这本质上是因为不再假设存在复制映射；在量子情形下，这是著名的“不可克隆”定理所致。

尽管经典因果模型以及上述量子因果模型的特殊子类都是这类基于有向无环图的模型的特例，但后者的适用范围严格更广——其具体实例，无论是经典的还是量子的，通常都不是因果模型。然而，它们以有趣的方式支持对某些因果查询进行定义明确的推广。

开放查询。现在，我们终于可以将抽象 Do 查询推广如下。给定 DAG G 的组合模型 $\mathbb{M}$ ，对于非输入顶点的任意子集 S，我们定义查询 $\mathrm{Open}(\mathsf{S})$，其表示仍以图形方式绘于下方左侧。$\mathrm{Open}(\mathsf{S})$ 的图表通过修改 $\mathsf{D}(G,\mathsf{V}^{\text{out}})$ 得到，具体方法是

[42]对于（经典）因果模型而言，输出 V^{out} 的选择是“无关紧要的”，因为任何改变都只是多“复制输出”一些变量或少复制一些变量，并不会改变因果机制；而定义 59 中的模型则可以在不存在复制映射的情况下定义（如量子理论中那样），因此输出 V^{out} 的选择就成了一个不可忽视的方面。[43]虽然我们可以直接将 U_j 视为通道 m_{X_j}，但以形式化方式引入 X_j ，能明确说明我们如何强调将 U_j 视为单个变量时的输出。

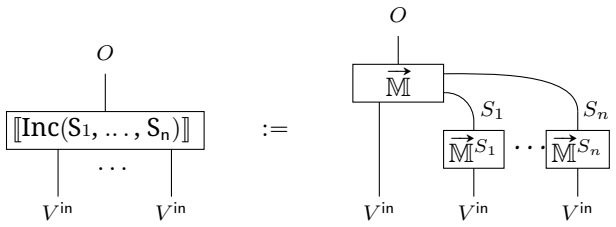
对于所有 X ∈S，删除盒子 c_X 以及所有导线 $e_{Y,\ X}$，其中 Y ∈ Pa(X)，然后将 X 设为图表的一个输入（见下方右侧）。[44]



交换查询。我们也可以相对于 D(G, V^{out}) 如下定义交换查询。若顶点的一个子集 W ⊆ V 满足：对于任意 X, Y ∈ W，若 X ≠ Y，则在 G 中不存在从 X 到 Y 的路径，我们就称 W 为可平行化的 。在这种情况下，对于某些图表 f、g 以及额外的导线 W′，我们可以将图表 D(G, V^{out}) 分解为下方左侧所示的形式。[45] 接下来，我们将右下方所示的对象定义为一个通道 $\vec{\mathbb{M}}^W$。



最后，对于任意一组两两不相交的子集 S1, . . . , Sn，其中每个子集都是可平行化的；并且对于输出的任意子集 $O \subseteq \overline{V^{\text{out}}}$，我们定义一个交换查询，其构造方式与式 (12) 完全相同，即：[46]

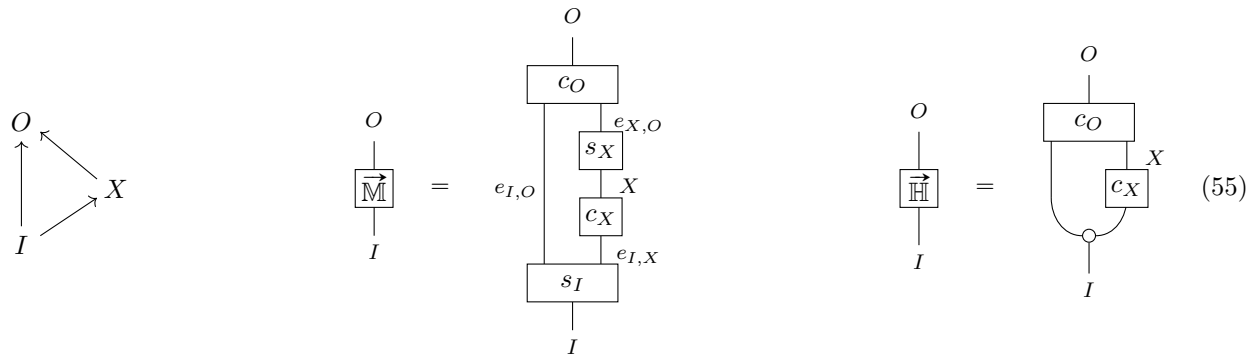


因此，开放操作查询和交换查询提供了两种查询开放有向无环图组合模型的方式，不局限于输入输出行为；输入输出行为可能出现在相应的向下抽象关系中。这

生成 lisation i一种有前景的 isi途径， f或用于说 l明目的，来 i绕过 t he subtleties 周围 d 因果模型的推广，并用两个例子说明其用途；在这两个例子中，我们通过抽象到一个因果模型，解释给定量子电路模型的若干方面（在例 58 和例 61 的意义上）。

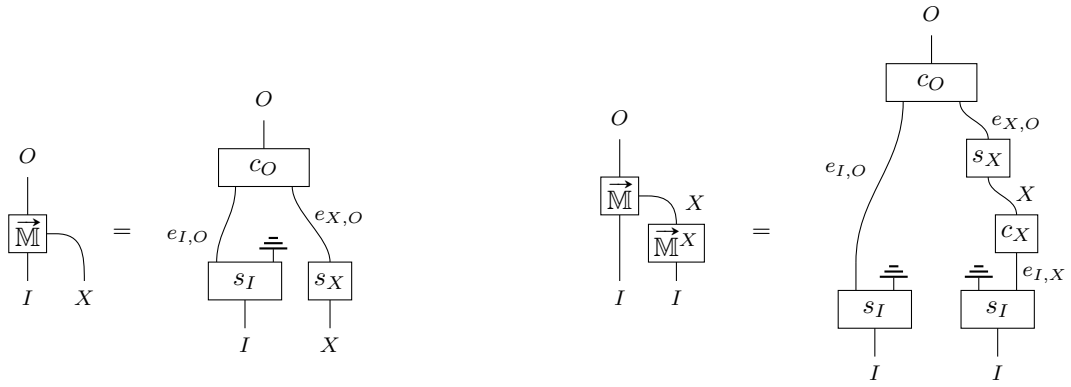
示例 63（在量子电路模型中解释线路）。假设我们想在量子电路模型中，将一组量子线路（例如量子比特）解释为对应于某个高层级的经典变量 X，该模型的输入为 I_L，输出为 O_L 𝕃 。为此，我们首先给出一个经典因果模型 $\mathbb{H}$，其中包含 X，其结构由下方左侧所示的简单有向无环图 G 表示。一个一般的组合模型 $\mathbb{M}$ 的 G 将由下方中央的图表D(G, V^{out}) 表示，而（经典）因果模型 $\mathbb{H}$ 则特化为右侧的形式。

[44]至于抽象 Do 查询，我们可以将开放操作视为到一个新模型的变换。对于任意非输入 S，通过将 S 标记为输入并删除其所有入边来定义开放有向无环图 $G_{\text{Open}(S)}$。D(G、V^{out}) 的任意模型 $\mathbb{M}$ 都会诱导出 $\mathcal{S}_{\text{D}(G_{\text{Open}(S)}, \text{V输出}\ )}$ 的一个模型 $\mathbb{M}' = \mathbb{M}_{\text{Open}(S)}$，并带有 ⟦开放 ($S$)⟧ = $\overrightarrow{\mathbb{M}_{\text{Open}(S)}}$；其定义方式是将 $[\![s_Y]\!]_{\mathbb{M}'}$ 定义为 ⟦s_Y⟧ 在 Ch(Y)\$e_{Y,X}$ 上的边缘分布，同时丢弃其 $e_{X,Y}$ 输出。[45] 注意，对于每个 X ∈ W，盒子 c_X 随后包含于 f 中，而 s_X 包含于 g 中。进而可以证明，用于定义 $\vec{\mathbb{M}}^W$ 的 f 的边缘分布确实独立于所选取的（非唯一的）f、g 及子集 W′。[46]需要考虑可并行化集合 W及相应的通道 $\vec{\mathbb{M}}^W$，这与针对（经典）因果模型的交换查询形成对比，并反映了一般情形下“将变量变为输出”这一非平凡步骤，而这一点前文已经强调过。

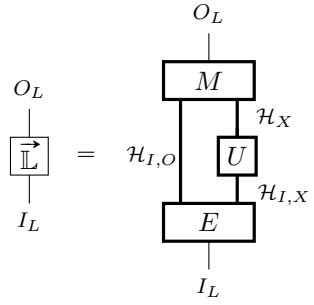


(55)

对于任何此类模型 型 $\mathbb{M}$ 上 G 关于在 X 处进行开放操作和交换的查询分别如下。



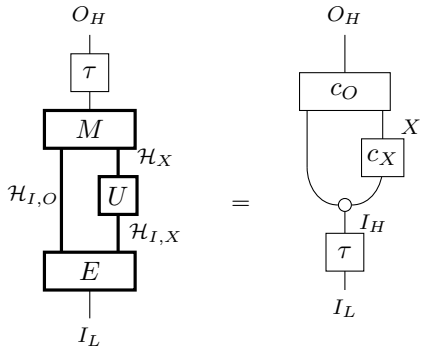
假设量子模型 $\mathbb{L}$由一个包含编码器、幺正变换和测量的电路图指定。我们必须首先将 $\mathbb{L}$ 视为 G的模型，方法是将这些通道组合成复合项；形式上，这是示例 *58* 中所示的严格组件级抽象（图中以虚线盒子表示）。完成这一步后，我们可以如上所述将 $\mathbb{L}$ 视为 G 的模型，并按下图进行因子分解。



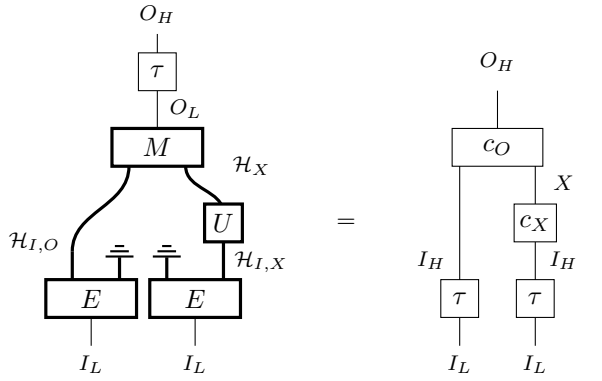
我们现在已将 $\mathbb{L}$ 和 $\mathbb{H}$ 都描述为 G 的模型，因此具有相同的签名，并试图根据 $\mathbb{H}$ 的经典变量 X 来解释幺正算子 U 的输出。为此，我们必须针对某些查询的选择展示一个抽象 $\mathbb{L} \to \mathbb{H}$ 。现在我们考虑三种形式的查询，以及相应的不同强性的抽象。

首先，仅在输入输出查询层级进行抽象，只需要分别在输入和输出上使用经典信道 $\tau\colon\ \ I_L \to I_H$ 和 $\tau\colon\ \ O_L \to O_H$ ，并满足以下条件，以确保相同

输入输出行为。

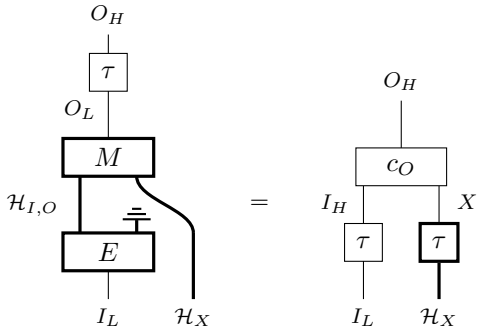


要直接指代 X，我们还应考虑进一步的查询。其次，基于交换查询的抽象需要满足一个额外条件，该条件对应于在 X 处进行交换。也就是说，插入第二个输入，并且仅在 $\mathcal{H}_X$ *(*左端*)*使用其嵌入，所得输出与根据该第二个输入确定 X （右端）时的输出相同。



(56)

最后，更强的要求是基于开放查询的抽象。这要求存在一个通道τ： $\mathcal{H}_X \to X$ （在实践中通过测量实现），满足在 X处进行开放操作的一致性条件。这将交换条件强化为适用于 $\mathcal{H}_X$ 的任意状态与 X的状态之间的条件，而不只是适用于通过编码器从输入得到的状态。

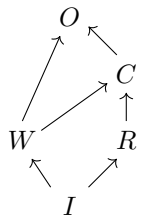


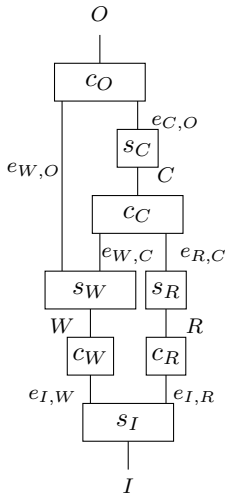
(57)

直观地说，(56)或(57) 都可以作为量子模型的附加条件，用来证明将 U 的输出 $\mathcal{H}_X$ 与变量 X联系起来是合理的。

示例 64（量子地理猜测器）。考虑一个量子模型 $\mathbb{L}$，它根据输入图像 I 给出对其位置的猜测 O。假设我们希望先评估天气W和图像中的道路类型 R，再根据这些信息判断国家 C 并得出最终输出 O，以检验该模型是否有效。我们可以将其表示为

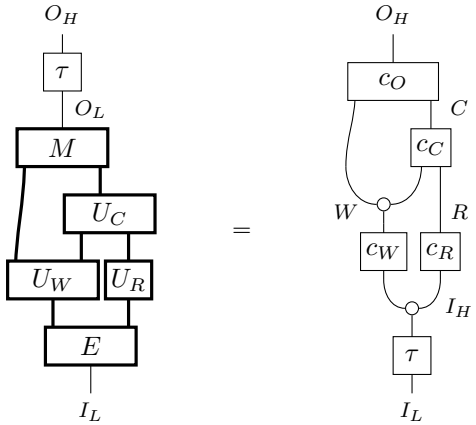
以下有向无环图 G 和模型结构 $\mathsf{D}(G,\ O)$所提出的假设。





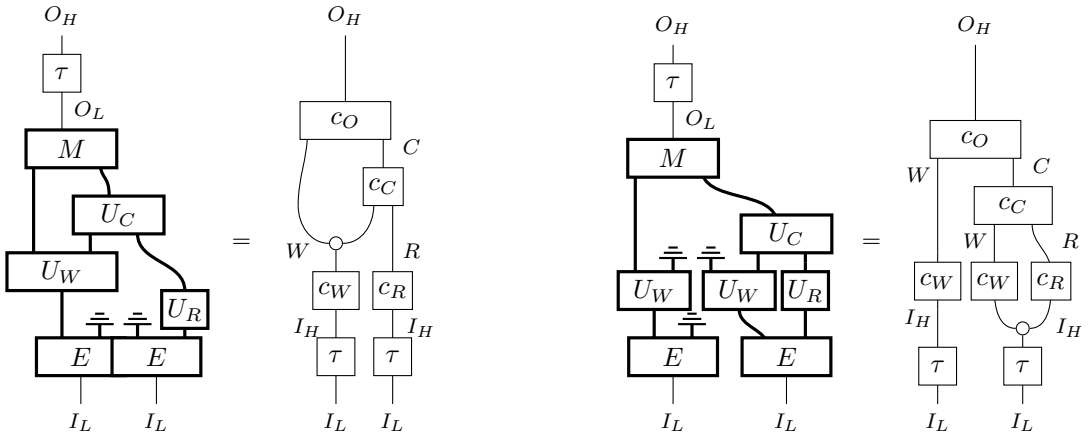
首先，必须能够将量子模型 $\mathbb{L}$ 的门分解为G模型的形式，类似于示例 *61*，并展示一个 G的经典因果模型 $\mathbb{H}$（在该模型中，我们可以将 R 的状态直接解释为道路，将 C 解释为国家，等等）。

现在，为了确定 $\mathbb{L}$ 在将 U_X的输出 $\mathcal{H}_X$ 与 X对应起来时，是否符合 $\mathbb{H}$，对于 $\mathbb{H}$的每个变量 X，我们必须针对某种查询选择给出一个抽象 $\mathbb{L} \to \mathbb{H}$ 。我们再次考虑三种可能性。首先，如果只考虑输入输出查询，就意味着这些模型在输入 $I_L \to I_H$ 和输出 $O_L \to O_H$上经由经典信道后，具有相同的输入输出行为，这相当于以下内容。

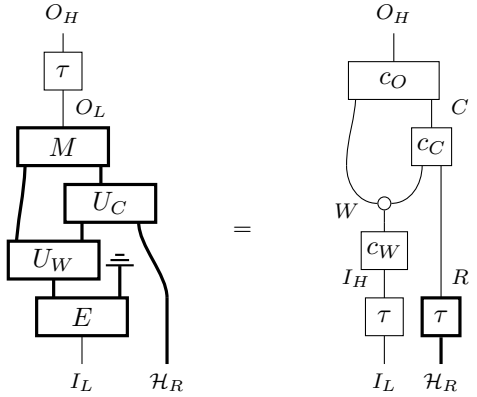


为了真正关联空间 $\mathcal{H}_R$ 与 R等，我们应该进一步采用至少基于交换查询的抽象。为此，我们为每个可并行化子集的顶点得出一个条件。$\{R\}$ 的条件如下所示，左侧的条件与 $\{W\}$ 的条件相同（只需交换等式两边的左、右输入），而 $\{C\}$ 的条件则在右侧。其他子集要么不可并行化，要么

导致平凡　条件，因为它们覆盖了上方相应图表的整个水平切片　　.



最后，基于开放查询的更强抽象将为每个变量 X包含一个通道 τ： $\mathcal{H}_X \to X$，并满足每个非输入变量子集的一致性条件（其中有些条件可能确实是平凡的）。例如，在 $\{R\}$ 处进行开放操作的一致性条件如下。



8 讨论

最后，我们讨论几点问题，并展望未来工作方向。

向下抽象与向上抽象。 这项工作的一个关键发现是，抽象关系自然分为两种形式上截然不同的类型：向下抽象（查询从高层映射到低层）和向上抽象（查询则沿相反方向，从低层映射到高层）。据我们所知，因果抽象领域的文献完全采用向上抽象，它在特定“具体”（Do-）干预层面上刻画因果抽象。

然而，我们的分析表明，向下抽象这一概念或许最为自然、最为根本。首先，向下抽象的形式描述很简单，由一个函子和一个自然变换构成；范畴论学家或许会很自然地将二者打包为模型之间的一个“态射”（见下文及附录 F）。其次，大多数更具体的因果抽象概念，更不用说实际例子，都属于向下抽象类型。除了一般性的精确变换（不附加其他假设），以及同构构造性抽象和同构交换抽象等明确的分布式因果抽象之外，我们所知的其他任何值得注意的概念，本质上都可视为向下抽象。后一类包括构造性抽象、交换抽象和反事实抽象；它们通常被表述为向上抽象，并引用具体干预，而这种表述是通过命题18从向下抽象得出的。

因此，尽管向上抽象的概念对于某些情形可能很重要，但它之所以受到重视，或许仅仅是因为人们关注具体干预，而对开放模型不够熟悉。相比之下，向下抽象或许会成为未来研究中更有帮助的默认抽象概念。

机制级因果抽象。我们的方法带来的一项新贡献是组件级抽象的概念，尤其是它的一种实例，即机制级因果抽象，这是一种新颖的、强化的构造性因果抽象形式。根据我们的定义，它是一种构造性抽象，并为每个高层因果机制补充了一致性条件。借助定理 51，我们看到，这实际上等价于两种表述：一方面，它等价于一种构造性抽象，其中低层变量的划分 π 满足某个纯粹的图论条件；另一方面，它等价于一个函子π^S：$\mathbf{S}_H \to \mathbf{S}_L$，以及一个与模型的整体网络图相协调的自然变换。值得注意的是，后一种表述意味着，无须直接提及 Do 查询（甚至不必提及我们所说的抽象 Do 查询），也无须提及任何干预，就能定义机制级因果抽象。

尽管它比构造性抽象严格更强，但它涵盖了构造性抽象的一些特殊情形，而这些情形似乎符合常见直觉：也就是说，在这些情形中，人们会选取给定（低层级）因果模型的某种变量划分，以相对直接的方式定义一个高层级因果模型，使二者构成构造性抽象。事实上，例如 [BH19] 中的构造性抽象主要都是机制级的（例 27 就是其中之一）。因此，未来值得从实践角度更深入地理解这一新颖概念：大多数实际示例是否实际上都属于这种强机制级类型？我们是否应当期待任何“良好”的构造性抽象实例都具备这一性质？

同构构造性抽象与强因果抽象。受 [GWP+24、GIZ+23] 的工作启发，我们定义了同构构造性抽象这一概念——它是由“同构诱导”的抽象与构造性抽象复合而成。这类抽象构成了一类基于一般（非 Do）干预的精确变换，其特点是高层变量具有分布式的低层级“表示”，且具备许多优良性质，同时不要求变量对齐是不相交的。除此之外，我们还以强因果抽象之名，概述了一些自然条件；人们或许会希望任何实用的因果抽象都满足这些条件。后者是构造性抽象的推广，与文献中的一些概念密切相关，但并不完全相同，主要区别在于它允许一般干预（见 C 节）。

现在，同构构造性抽象也推广了构造性抽象，并且确实是强因果抽象的一个实例。因此，一个自然的问题是：强因果抽象还有哪些（如果有的话）并非同构构造性抽象的例子？

我们注意到，能够进一步将同构构造性抽象中的构造性情形单独区分出来的理解条件（例如仅限于 Do-干预），与 [BH19] 中提出的一个猜想有关。该猜想涉及他们所谓的强 τ-抽象与构造性 τ-抽象之间的等价性（另见 C 节）。此外，为了更好地理解具有不相交变量对齐的抽象这一领域，若能在我们的框架中阐明 [XB24] 中抽象不变性条件被违反的意义，以及 [XB25], 中投影抽象的组合性质，还有图形一致性、簇有向无环图和 [SZ24] 中的结果所代表的意义，将会很有趣。

具有代数结构的查询抽象。虽然我们对抽象的主要定义仅涉及组织成签名 $\mathcal{Q}$ 的查询集合，但后者可能具有进一步的非平凡代数结构。我们定义了 $\mathcal{Q}$ 具有幺半群结构意味着什么，以及向上抽象的映射 ω 成为保持查询代数结构的同态意味着什么（B 节）。这使我们能够将因果模型文献中讨论的“保序”视为一种特殊情形，并表明要求更强的同态性质或许是更自然的条件。未来若能进一步对查询上的这一额外复合层次作范畴论处理，将很值得探讨。

量子人工智能的抽象。在第 7 节中，我们介绍了量子电路与高层经典因果模型之间的抽象。我们推测，这些抽象或许能够以有原则的方式约束量子模型，使其按照高层经典因果模型的行为方式运行，从而能够根据该模型对其进行解释。从一开始就施加此类约束（例如在训练过程中），或许能为获得可解释的

量子人工智能模型提供一种途径，正如训练神经网络使其按照因果模型的行为方式运行时，也可以应用因果抽象（另见第 1 节）。

如果是这样，一个重要的问题是，在这种抽象中应采用哪些查询，而不仅仅是匹配输入输出行为。我们在例63和例64中展示了与不同查询相对应的一致性条件。交换查询或许最为实用，因为只需在经典输入上进行测试即可。然而，由于这类查询只能针对可并行化子集执行，因此可能不足以完整捕捉某些高层结构（有向无环图）。这种方法的另一个问题是，在某些情况下，抽象施加的约束可能会去除量子模型中必要的相干性，从而从根本上削弱其用途。

未来，探讨这些问题并更全面地评估抽象作为一种为量子模型赋予可解释结构的手段的潜在价值，将会很有意义。

连续参数化量子电路的抽象。虽然我们在第 7 节中只研究了有限经典系统，但在实践中，量子模型通常由参数化量子电路实现，其输入 I 和输出 O 取值连续，位于 $\mathbb{R}^n$。此外，参数化量子电路所描述的整体函数通常通过对测量（或一组测量）取期望值来定义，从而得到一个确定性映射 $I \to O$，将每个输入映射到相应的期望值。事实上，可以将**量子-经典**升级为一个以类似方式定义的范畴**混合范畴**，其中包含诸如 $\mathbb{R}$ 之类的无限经典系统，并将期望值作为**混合范畴**经典子范畴上的额外图形元素来建模 [Tul26]。未来，若能扩展我们的方法，使其涵盖作用于此类参数化量子电路的抽象，将会很有意义。

量子因果抽象。 第 7 节中，我们探讨了从量子电路（视为组合模型）到高层经典因果模型的抽象；但若要描述一种真正因果的、从量子到经典的抽象，则需要从量子因果模型 [BLO19, ABH$^+$17]出发进行抽象。Costa 和 Shrapnel 提出了一类本质上属于量子因果模型的子类，这类模型可以定义为组合模型（例 62） [CS16, Gia19], ，其定义本质上依据的是组合结构，而非因果结构。然而，一般而言，量子因果模型是通过“高阶映射”来定义的，因此目前尚无法直接将其视为组合模型（这涉及尚未解决的技术问题 [vdLL25, LB21]）。尽管如此，从量子理论基础和因果哲学的角度来看，量子因果抽象的概念仍值得探讨，尤其是可将其作为研究退相干何时可能促成“与因果相一致的经典实在涌现”的一种途径。[47]

因此，真正的量子因果抽象仍有待未来工作来发展；第 7 节和附录 D 则提供了可能的起点：首先扩展组合模型的概念，以纳入高阶映射。

循环结构与近似版本。本文所采用的因果模型概念——尽管在许多方面比不少文献中的概念更为一般——确实假定了无环性。然而，有理由基于一般而言存在循环的模型来研究因果抽象。例如，从因果模型的视角研究随时间演化的系统的平衡化 [RWB$^+$17]，以及有人认为，可解释人工智能领域对机器学习模型的可解释性研究也促使人们考虑循环性 [GIZ$^+$23]。[48] 将因果模型与抽象的范畴论处理推广到循环因果模型，将会很有意思。一种自然的做法是通过由模型 $\mathbb{M}$ 诱导的并行机制过程 $\mathcal{F}_{\mathbb{M}}$，结合附录 D 中概述的“迹技巧”来实现；另一个可能的起点是 [FGV25] 的框架。这两种方法也都自然地与量子因果模型抽象的研究相联系。

[47]此外，还有其他量子因果模型框架，但它们似乎并不适合研究此处提出的基础性问题。除了 [BLO19、ABH$^+$17] 的框架以及 [CS16, Gia19], 中的特殊情形之外，Henson、Lal 和 Pusey [HLP14] 提出的概念也并非“完全量子化”，因为经典变量（例如测量结果）已被内置于该概念本身；其他具有这一特点的框架还包括 [Fri12, Las07, PB15, RAV$^+$15, FGV25]；关于这些框架及其他框架的综述，参见 [Lor20, ABH$^+$17] 。[48]不过请注意，与循环因果模型领域的其他文献不同， [GIZ$^+$23] 中的概念是可能的最弱概念——本质上，变量上的任意函数，只要不要求解唯一，也不要求其在外生变量上产生输出的良定义概率分布，就能赋予一定的随机性。允许比 Do- 干预更一般的干预，同时也允许循环结构，但排除悖论模型或不一致模型，是一项非平凡的任务，迄今尚无简单的刻画方式。

还可以尝试沿另一个方向扩展抽象的范畴论处理，即允许采用 [BEH20, GWP+24, GIZ+23] 所研究意义上的近似抽象，这对于实际机器学习场景十分重要。在这种情况下，不再要求每个查询的一致性条件严格成立，而是要求相应的左侧与右侧在某种适当意义上彼此接近。

进一步形式化。从范畴论视角来看，进一步形式化本文的核心定义也是一个值得考虑的自然步骤。其基础是定义一个范畴 **模型**(**C**) ，其对象是 **C** 中的模型，而态射则实质上同时包含向下抽象和向上抽象。有关这一方法的更多细节和基本定义，请参见附录 F；未来若能对其进行全面发展，将会很有意义。

利用我们的方法探究（因果）抽象与计算机科学中的抽象范畴概念（例如编程语言或类型之间的抽象）之间的联系，也会很有意义。后者通常也通过函子和自然变换来形式化，而它们如今还构成了伴随。

参考文献

[ABH+17] John-Mark A Allen、Jonathan Barrett、Dominic C Horsman、Ciar´an M Lee 和 Robert W Spekkens。量子共同原因与量子因果模型。*Physical Review X*，7(3):031021，2017年。[BCII22] Elias Bareinboim、Juan D Correa、Duligur Ibeling 和 Thomas Icard。论 Pearl 层级与因果推断的基础。载于*Pearl 的概率与因果推断研究*，第507–556页。2022年。[BDHLC22] Johann Brehmer、Pim De Haan、Phillip Lippe 和 Taco S Cohen。弱监督因果表征学习。神经信息处理系统进展，35:38319–38331，2022年。[Bec22] Sander Beckers。因果解释与可解释人工智能。载于*Conference on Causal Learning and Reasoning*，第90–109页。PMLR，2022年。[BEH20] Sander Beckers、Frederick Eberhardt 和 Joseph Y Halpern。近似因果抽象。载于人工智能不确定性会议，第606–615页。PMLR，2020年。[BH19] Sander Beckers 和 Joseph Y Halpern。抽象因果模型。载于*AAAI 人工智能会议论文集*，第33卷，第2678–2685页，2019年。[BLO19] Jonathan Barrett、Robin Lorenz 和 Ognyan Oreshkov。量子因果模型。*arXiv 预印本 arXiv:1906.10726*，2019年。[CB20] Juan Correa 和 Elias Bareinboim。随机干预演算：因果效应识别与替代实验。载于*AAAI 人工智能会议论文集*，第34卷，第10093–10100页，2020年。[CEP16] Krzysztof Chalupka、Frederick Eberhardt 和 Pietro Perona。多层次因果效应系统。载于人工智能与统计，第361–369页。PMLR，2016年。[CEP17] Krzysztof Chalupka、Frederick Eberhardt 和 Pietro Perona。因果特征学习：综述。*Behaviormetrika*，44(1):137–164，2017年。[CJ19] Kenta Cho 和 Bart Jacobs。通过弦图实现分解与贝叶斯逆变换。《计算机科学中的数学结构》，29(7):938–971，2019年。

[CK18] Bob Coecke 和 Aleks Kissinger。量子过程图示：量子理论与图示推理入门课程。载于《图示表示与推理：第十届国际会议，Diagrams 2018，爱丁堡，英国，2018年6月18日至22日，论文集》第10卷，第28–31页。施普林格，2018年。[Coe06] Bob Coecke。向执业物理学家介绍范畴。载于《什么是范畴论》，第30卷，第45–74页。Polimetrica Publishing Milan，2006年。[Coe08] Bob Coecke。Selinger 的 CPM 构造对混合态的公理化描述。《理论计算机科学电子札记》，210:3–13，2008年。[Coe14] Bob Coecke。终端性蕴含非信号传递。arXiv 预印本 arXiv:1405.3681，2014年。[Coe21] Bob Coecke。文本结构的数学。载于《Joachim Lambek：数学、逻辑与语言学的交织》，第181–217页，2021年。[CS16] Fabio Costa 和 Sally Shrapnel。量子因果建模。《新物理学杂志》，18(6):063032，2016 年。[CSC10] Bob Coecke、Mehrnoosh Sadrzadeh 和 Stephen Clark。组合分布式意义模型的数学基础。《语言分析》，36(1):345–384，2010年。[DB25] Gabriele D'Acunto 和 Claudio Battiloro。因果知识的相对性。arXiv 预印本 arXiv:2503.11718，2025年。[DDLB26] Gabriele D'Acunto、Paolo Di Lorenzo 和 Sergio Barbarossa。学习一致的因果抽象网络。arXiv 预印本 arXiv:2602.02623，2026年。[dF22] Giovanni de Felice。自然语言处理的范畴工具。arXiv 预印本 arXiv:2212.06636，2022年。[dFTC20] Giovanni de Felice、Alexis Toumi 和 Bob Coecke。Discopy：Python 中的幺半范畴。arXiv 预印本 arXiv:2005.02975，2020年。[DLB25] Gabriele D'Acunto、Paolo Di Lorenzo 和 Sergio Barbarossa。因果抽象网络：层论框架。2025年。[DZFDL25] Gabriele D'Acunto、Fabio Massimo Zennaro、Yorgos Felekis 和 Paolo Di Lorenzo。基于语义嵌入原理的因果抽象学习。arXiv 预印本 arXiv:2502.00407，2025年。[ED25] Markus Englberger 和 Devendra Singh Dhami。因果抽象的范畴统一。arXiv 预印本 arXiv:2510.05033，2025年。[FGV25] Carla Ferradini、Victor Gitton 和 V Vilasini。基于图分离定理的循环量子因果建模。arXiv 预印本 arXiv:2502.04168，2025年。[FK23] Tobias Fritz 和 Andreas Klingler。范畴概率中的 d-分离判据。《机器学习研究杂志》，24(46):1–49，2023年。[FL23] Tobias Fritz 和 Wendong Liang。自由 gs-幺半范畴与自由马尔可夫范畴。《应用范畴结构》，31(2):21，2023年。[Fri12] Tobias Fritz。超越贝尔定理：关联情景。《新物理学杂志》，14(10):103001，2012年。[Fri20] Tobias Fritz。马尔可夫核、条件独立与充分统计量定理的综合方法。《数学进展》，370:107239，2020年。[GGM+24] Tarun Gupta、Wenbo Gong、Chao Ma、Nick Pawlowski、Agrin Hilmkil、Meyer Scetbon、Marc Rigter、Ade Famoti、Ashley Juan Llorens、Jianfeng Gao 等。因果性对于具身人工智能基础世界模型的关键作用。arXiv 预印本 arXiv:2402.06665，2024年。

[GHI25] Atticus Geiger、Jacqueline Harding 和 Thomas Icard。因果抽象如何支撑计算解释。arXiv 预印本 arXiv:2508.11214，2025年。[Gia19] Christina Giarmatzi。一种量子因果发现算法。载于《重新思考量子力学中的因果论》，第125–150页。施普林格，2019年。[GIZ+23] Atticus Geiger、Duligur Ibeling、Amir Zur、Maheep Chaudhary、Sonakshi Chauhan、Jing Huang、Aryaman Arora、Zhengxuan Wu、Noah Goodman、Christopher Potts 等。因果抽象：机制可解释性的理论基础。arXiv 预印本 arXiv:2301.04709，2023年。[GLIP21] Atticus Geiger、Hanson Lu、Thomas Icard 和 Christopher Potts。神经网络的因果抽象。《神经信息处理系统进展》，34：9574–9586，2021年。[GRP20] Atticus Geiger、Kyle Richardson 和 Christopher Potts。神经自然语言推理模型部分嵌入了词汇蕴涵与否定理论。载于 Afra Alishahi、Yonatan Belinkov、Grzegorz Chrupa-la、Dieuwke Hupkes、Yuval Pinter 和 Hassan Sajjad 编，《第三届用于分析和解释自然语言处理神经网络的 BlackboxNLP 研讨会论文集》，第163–173页，线上，2020年11月。计算语言学协会。[GWL+22] Atticus Geiger、Zhengxuan Wu、Hanson Lu、Josh Rozner、Elisa Kreiss、Thomas Icard、Noah Goodman 和 Christopher Potts。为可解释神经网络诱导因果结构。载于《国际机器学习会议》，第7324–7338页。PMLR，2022年。[GWP+24] Atticus Geiger、Zhengxuan Wu、Christopher Potts、Thomas Icard 和 Noah Goodman。寻找可解释因果变量与分布式神经表征之间的对齐。载于《因果学习与推理》，第160–187页。PMLR，2024年。[HLP14] Joe Henson、Raymond Lal 和 Matthew F Pusey。广义贝叶斯网络中相关性的理论无关界限。《新物理学杂志》，16(11)：113043，2014年。[HLT25] Jingzhou Huang、Jiuyao Lu 和 Alexander Williams Tolbert。社会科学中的因果特征学习。arXiv 预印本 arXiv:2503.12784，2025年。[HT22] Yaojie Hu 和 Jin Tian。神经元依赖图：神经网络的一种因果抽象。载于《国际机器学习会议》，第9020–9040页。PMLR，2022年。[JKZ19] Bart Jacobs、Aleks Kissinger 和 Fabio Zanasi。通过弦图手术进行因果推断。载于《软件科学与计算结构基础：第22届国际会议 FOSSACS 2019》，作为《欧洲软件理论与实践联合会议》的一部分在捷克共和国布拉格举行，2019年4月6日至11日，论文集22，第313–329页。施普林格，2019年。[KFY+21] Dimitri Kartsaklis、Ian Fan、Richie Yeung、Anna Pearson、Robin Lorenz、Alexis Toumi、Giovanni de Felice、Konstantinos Meichanetzidis、Stephen Clark 和 Bob Coecke。lambeq：用于量子自然语言处理的高效高层级 Python 库。arXiv 预印本 arXiv:2110.04236，2021年。[KGSL] Avinash Kori、Ben Glocker、Bernhard Sch¨olkopf 和 Francesco Locatello。统一因果表征学习与以对象为中心的表征学习，从而实现因果组合。载于《ICLR 2025 世界模型研讨会：理解、建模与扩展》。[KSB23] Armin Keki´c、Bernhard Sch¨olkopf 和 Michel Besserve。因果模型的定向约简。arXiv 预印本 arXiv:2311.18639，2023年。[Las07] Kathryn Blackmond Laskey。量子因果网络。载于《AAAI 春季研讨会：量子交互》，第142–149页，2007年。[LB21] Robin Lorenz 和 Jonathan Barrett。幺正变换的因果与组合结构。《Quantum》，5：511，2021年。[Lei14] Tom Leinster。《基础范畴论》，第143卷。剑桥大学出版社，2014年。

[LKR25] Xiusi Li、S´ekou-Oumar Kaba 和 Siamak Ravanbakhsh。论因果抽象的可识别性。arXiv 预印本 arXiv:2503.10834，2025年。[Lor20] Robin Lorenz。量子因果结构。博士论文，牛津大学，2020年。[LT23] Robin Lorenz 和 Sean Tull。弦图中的因果模型。arXiv:2304.07638，2023年。[MGIB23] Riccardo Massidda、Atticus Geiger、Thomas Icard 和 Davide Bacciu。采用软干预的因果抽象。载于《因果学习与推理会议》，第68–87页。PMLR，2023年。[ML13] 桑德斯·麦克莱恩。《工作数学家的范畴论》，第5卷。施普林格科学与商业媒体，2013年。[MSEI25] Milan Moss´e、Kara Schechtman、Frederick Eberhardt 和 Thomas Icard。利用因果抽象建模歧视。arXiv 预印本 arXiv:2501.08429，2025年。[OS22] Jun Otsuka 和 Hayato Saigo。论因果模型的等价性：一种范畴论方法。载于《因果学习与推理会议》，第634–646页。PMLR，2022年。[PB15] Jacques Pienaar 和 ˇCaslav Brukner。量子因果模型的图分离定理。《新物理学杂志》，17(7)：073020，2015年。[Pea09] Judea Pearl。《因果论》。剑桥大学出版社，2009年。[PMG25] Theodora-Mara Pˆıslar、Sara Magliacane 和 Atticus Geiger。组合因果模型，以获得更准确的神经网络抽象。arXiv 预印本 arXiv:2503.11429，2025年。[PZ25] Robin Piedeleu 和 Fabio Zanasi。《计算机科学家的弦图导论》。应用范畴论读本。剑桥大学出版社，2025年。[RAV+15] Katja Ried、Megan Agnew、Lydia Vermeyden、Dominik Janzing、Robert W Spekkens 和 Kevin J Resch。推断因果结构的量子优势。《自然 · 物理学》，11(5)：414–420，2015 年。[RE24] Jonathan Richens 和 Tom Everitt。稳健智能体会学习因果世界模型。载于《第十二届国际学习表征会议》，2024年。[REA25] Jonathan Richens、Tom Everitt 和 David Abel。通用智能体需要世界模型。载于《第四十二届国际机器学习会议》，2025年。[RERS23] Thomas S Richardson、Robin J Evans、James M Robins 和 Ilya Shpitser。无环有向混合图的嵌套马尔可夫性质。《统计学年刊》，51(1)：334–361，2023年。[Ris20] Eigil Fjeldgren Rischel。因果模型的范畴论。硕士论文，哥本哈根大学，2020年。[RW21] Eigil F Rischel 和 Sebastian Weichwald。组合抽象误差与因果模型范畴。载于《人工智能不确定性会议》，第1013–1023页。PMLR，2021年。[RWB+17] Paul K Rubenstein、Sebastian Weichwald、Stephan Bongers、Joris M Mooij、Dominik Janzing、Moritz Grosse-Wentrup 和 Bernhard Sch¨olkopf。结构方程模型的因果一致性。arXiv 预印本 arXiv:1707.00819，2017年。[Sel10] Peter Selinger。幺半范畴图形语言综述。载于《物理学新结构》，第289–355页。施普林格，2010年。[Sel11] Peter Selinger。幺半范畴图形语言综述。《物理学新结构》，第289–355页，2011年。[SLB+21] Bernhard Sch¨olkopf、Francesco Locatello、Stefan Bauer、Nan Rosemary Ke、Nal Kalchbrenner、Anirudh Goyal 和约书亚·本吉奥。迈向因果表征学习。《IEEE会刊》，109(5)：612–634，2021年。

[SZ24] Willem Schooltink 和 Fabio Massimo Zennaro。图形因果抽象与函数因果抽象的对齐。arXiv 预印本 arXiv:2412.17080，2024年。[TLC+24] Sean Tull、Robin Lorenz、Stephen Clark、Ilyas Khan 和 Bob Coecke。迈向面向可解释人工智能的组合式可解释性。arXiv 预印本 arXiv:2406.17583，2024年。[TO26] Sean Tull 和 Masanao Ozawa。类量子认知的组合分析（即将发表）。2026年。[Tul26] Sean Tull。用于量子-经典混合计算的形式图（即将发表）。2026年。[vdLL25] Tein van der Lugt 和 Robin Lorenz。酉因果分解：基于格理论的组合刻画，2025年。[WGH+24] Zhengxuan Wu、Atticus Geiger、Jing Huang、Aryaman Arora、Thomas Icard、Christopher Potts 和 Noah D Goodman。对 Makelov 等人（2023年）“可解释性幻觉”论点的回应。arXiv 预印本 arXiv:2401.12631，2024年。[WGI+24] Zhengxuan Wu、Atticus Geiger、Thomas Icard、Christopher Potts 和 Noah Goodman。规模化可解释性：识别 Alpaca 中的因果机制。《神经信息处理系统进展》，第36卷，2024年。[WMLC23] Vincent Wang-Mascianica、Jonathon Liu 和 Bob Coecke。将文本提炼为电路。arXiv 预印本 arXiv:2301.10595，2023年。[XB24] Kevin Xia 和 Elias Bareinboim。神经因果抽象。载于《AAAI 人工智能会议论文集》，第38卷，第20585–20595页，2024年。[XB25] Kevin Xia 和 Elias Bareinboim。有损表征下的因果抽象推断。arXiv 预印本 arXiv:2509.21607，2025年。[YRC+24] Dingling Yao、Dario Rancati、Riccardo Cadei、Marco Fumero 和 Francesco Locatello。以不变性原理统一因果表征学习。arXiv 预印本 arXiv:2409.02772，2024年。[YXL+23] Dingling Yao、Danru Xu、S´ebastien Lachapelle、Sara Magliacane、Perouz Taslakian、Georg Mar-tius、Julius von K¨ugelgen 和 Francesco Locatello。具有部分可观测性的多视图因果表征学习。arXiv 预印本 arXiv:2311.04056，2023年。[Zen22] Fabio Massimo Zennaro。结构因果模型之间的抽象：定义与性质综述。arXiv 预印本 arXiv:2207.08603，2022年。[Zha24] Yihao Zhang。模型可解释性中的因果抽象：简要综述。arXiv 预印本 arXiv:2410.20161，2024年。[ZMSB24] Yuchen Zhu、Sergio Hernan Garrido Mejia、Bernhard Sch¨olkopf 和 Michel Besserve。无监督因果抽象。载于 NeurIPS 2024 因果表征学习研讨会，2024年。

A 证明

注 *20*。 任何幺单函子 π: $\mathbf{V}_H^* \to \mathbf{V}_L^*$ 都会将 V 中的每个 V 映射到 V_H 中的某个列表 $\pi(\mathrm{V})$，并且必须保持恒等态射。由于 π 将查询映射到查询，因此对于 V_L中的某个子集 X，有 $\pi(Q_{(\mathrm{V})}) = Q_{\mathrm{X}}$，然后$\pi(V) = $X。但对于某个子集 Y， $\pi(Q_{\mathrm{V}_H}) = Q_{\mathrm{Y}}$ 必须成立，而该子集必须等于 V $\in \mathrm{V}_H$对应的 π(V) 的串接。因此， π(V) 必须两两不相交，因为 Y 不包含重复元素。 τ 的性质正是向下抽象的性质。 □

定理*25*。给定一个构造性抽象 $(\pi,\ \tau)$ 我们定义向下抽象如下。我们通过串接在类型上定义π，并在查询上定义为

$$\pi\left(\mathrm{Do}(\mathsf{S})\colon \mathsf{V}_{\mathsf{H}}^{\mathsf{in}}\otimes\mathsf{S}\to\mathsf{X}\right) := \left(\mathrm{Do}(\pi(\mathsf{S}))\colon \pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}})\otimes\pi(\mathsf{S})\to\pi(\mathsf{X})\right) \tag{58}$$

利用 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}}) = \mathsf{V}_{\mathsf{L}}^{\mathsf{in}}$。我们按照 (14) 中的方式，通过取乘积来定义任意类型上的 τ。对 (21) 取边缘分布，可得所有必要的自然性方程，精确地说明 (π,τ) 构成一个向下抽象。

反过来，设 (π,τ) 是一个向下抽象。由于 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}}) = \mathsf{V}_{\mathsf{L}}^{\mathsf{in}}$，且 π 是一个将查询映射为查询的幺半函子，因此它必定具有 (58) 的形式，因为右端是输入类型为 $\mathsf{V}_{\mathsf{L}}^{\mathsf{in}}\otimes\pi(\mathsf{S})$的唯一查询。现在考虑高层查询 Q，其输出为 $\mathsf{V}_{\mathsf{H}}^{\mathsf{out}}$，并对所有 $\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}}$ 进行干预。于是 $\pi(\mathsf{Q})$ 是低层查询，其输出为 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{out}})$，并对 $\pi(\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}})$进行干预。因此，$\pi(\mathsf{X})$ 全都是子集，且 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{out}}) \subseteq \mathsf{V}_{\mathsf{L}}^{\mathsf{out}}$。由于低层输入 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}} \cup \mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}})$ 没有重复项，$\pi(\mathsf{X})$ 必定两两不相交。因此，(π,τ) 是一个不相交变量对齐。自然性随即保证 (21) 成立。 □

命题 *29*。给定一个交换抽象，按照 (26) 在查询上定义 π，并按通常方式通过取乘积将 π、τ 扩展到任意类型。所有查询一致性条件都可通过取边缘分布由 (25) 推出。

反过来，设 $(\pi、\ \tau)$ 是这样的向下抽象。由于它必须将查询映射为查询，考察其输出可知，我们必须有 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{out}}) \subseteq \mathsf{V}_{\mathsf{L}}^{\mathsf{out}}$；根据假设，我们还有 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}}) = \mathsf{V}_{\mathsf{L}}^{\mathsf{in}}$ 和 $\pi(\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}}) \subseteq \mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{L}}$。因此，对于 X，$\pi(\mathsf{X})$ 必须全部是不相交子集，且这些子集位于 $\mathsf{V}_{\mathsf{H}}^{\mathsf{in}}$ 或 $\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{H}}$ 之内；又由于 $\mathsf{V}_{\mathsf{L}}^{\mathsf{in}}$ 和 $\mathsf{V}^{\mathsf{in}^\perp}{}_{\mathsf{L}}$ 不相交，所以这些子集也都彼此不相交。因此，$(\pi、\ \tau)$ 构成一个不相交变量对齐。然后由 τ 的自然性可得 (25)。 □

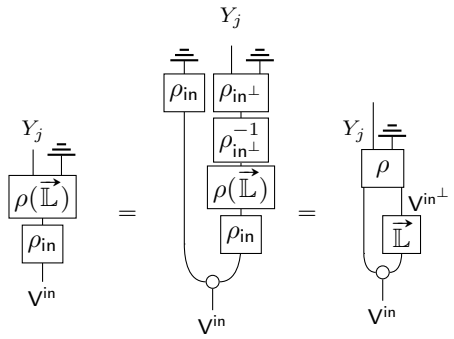
命题 *43*。根据假设，该映射定义良好；接下来只需验证自然性条件 (42)。

只需证明 $\rho_{\mathsf{in}^\perp}\circ\mathbb{M} = \rho(\mathbb{M})\circ\rho_{\mathsf{in}}$，因为这样一来，对于任意干预 σ，我们只需将 $\mathbb{M}$ 替换为 $\mathbb{M}_\sigma$，即可得到（42）。设 v_{in} 为 V^{in} 的一个尖锐状态。设 $v_{\mathsf{nin}} = \vec{\mathbb{M}}\circ v_{\mathsf{in}}$，并将 $v = v_{\mathsf{in}}\otimes v_{\mathsf{nin}}$ 定义为 V 的一个状态。由引理 40，我们有 $v = \mathcal{F}_{\mathbb{M}}\circ v$。因此 $\rho\circ v = \rho\circ\mathcal{F}_{\mathbb{M}}\circ v = \mathcal{F}_{\rho(\mathbb{M})_P}\circ\rho\circ v$。再次由引理 40，后者成立当且仅当，记 $\rho\circ v = w_{\mathsf{in}}\otimes w_{\mathsf{nin}}$，我们有 $w_{\mathsf{nin}} = \overrightarrow{\rho(\mathbb{M})}\circ w_{\mathsf{in}}$。由于 $\overrightarrow{\rho(\mathbb{M})}\circ\rho_{\mathsf{in}}\circ v_{\mathsf{in}} = \rho_{\mathsf{in}^\perp}\circ\vec{\mathbb{M}}\circ v_{\mathsf{in}}$如（41）那样分解为因子，我们有 $w_{\mathsf{in}} = \rho_{\mathsf{in}}\circ v_{\mathsf{in}}$ 和 $w_{\mathsf{nin}} = \rho_{\mathsf{in}^\perp}\circ v_{\mathsf{nin}} = \rho_{\mathsf{in}^\perp}\circ\vec{\mathbb{M}}\circ v_{\mathsf{in}}$。因此。由于 v_{in} 是任意的（并且假定 $\mathbf{C}$ 具有足够多的状态），我们按要求得出 $\overrightarrow{\rho(\mathbb{M})}\circ\rho_{\mathsf{in}} = \rho_{\mathsf{in}^\perp}\circ\vec{\mathbb{M}}$。 □

方程 (45)和 (46)。我们先验证 (45)。令 break_p 表示将状态 s 替换为 p、将 S 替换为 S'后 (39) 的左端。根据定义：

$$\mathcal{F}_{\mathsf{Ddo}(S'=p,\rho)} := \rho^{-1}\circ\mathcal{F}_{\rho(\mathbb{L})_{\mathrm{Do}(S'=p)}}\circ\rho = \rho^{-1}\circ\mathsf{break}_p\circ\mathcal{F}_{\rho(\mathbb{L})}\circ\rho = \rho^{-1}\circ\mathsf{break}_p\circ\rho\circ\mathcal{F}_{\mathbb{L}}$$

最后一步使用 (39)，然后使用 (42)。接下来，方程 (46) 是一个特殊情形，使用 (39) 的右端；注意到对于任意输出子集 Y_j（特别是对于 $Y_j = \pi(S_j)$），我们有：



最后一步再次使用 (42)。 □

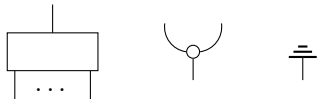
命题 *47* 的证明。将 $[-]_{\mathbb{L}}\circ\pi^S \Rightarrow [-]_{\mathbb{H}}$ 与 $1_{|-|_{\mathbb{H}}}$ 上的恒等变换 $|-|_{\mathbb{H}}$ 复合，我们得到一个变换 $[\![-]\!]_{\mathbb{L}}\circ\pi \Rightarrow [\![-]\!]_{\mathbb{H}}$：$\tau\circ 1_{|-|_{\mathbb{H}}}$，因为 $[-]_{\mathbb{L}}\circ\pi^S\circ|-|_{\mathbb{H}} = [-]_{\mathbb{L}}\circ|-|_{\mathbb{L}}\circ\pi = [\![-]\!]_{\mathbb{L}}\circ\pi$ 和 $[-]_{\mathbb{H}}\circ|-|_{\mathbb{H}} = [\![-]\!]_{\mathbb{H}}$。 □

A.1 机制层面的因果抽象

我们不我们致力于证明我们对机制级因果抽象的刻画。

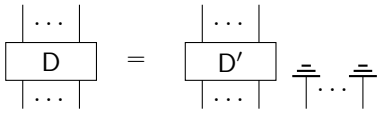
我们将使用以下内容。首先，回顾与因果模型 $\mathbb{M}$的开放有向无环图 $G = G_{\mathbb{M}}$相关联的网络图 $\mathrm{D}_{\mathbb{M}}$ 。我们可以如下刻画这类图表。

定义 65。 网络图 是一种由单输出盒子、复制映射和丢弃构成的字符串图D ：



此外，导线上带有标记；凡是未通过一串复制映射相连的导线，都赋予不同的标记，并且每个标记至多出现一次作为输出，且至多出现一次作为任意给定盒子的输入。

这类图表通过第 3.1 节前文所述的对应关系等价于开放有向无环图。若图表中的每个非输入变量都有一条通往输出的路径，我们就称网络图 D 为规范化的 。除输入线的交换外，任何此类图表都具有如下形式:



其中 D′ 是一个不包含 的网络图。

引理 66. 设 $\mathcal{S} = \mathcal{S}_{\mathbb{M}}$ 为开放因果模型 $\mathbb{M}$ 的签名，其输入变量为 V^{in} ，并设 X,Y为来自 $\mathcal{S}$的变量列表，其中每个列表都不包含重复变量。

1. $\mathcal{S}_{\mathrm{cd}} = \mathcal{CD}(\mathcal{S})$ 中的图表 $D : X \to Y$ 是规范化网络图，当且仅当对于每个使得 c_{V} 出现在 D 中的变量 V，我们有 c_{V} 仅出现一次，V 不是 D 的输入，且 V 在 D 中有一条通向输出的向上路径。*2.* 假设在 G 中存在一条不经过 X 的路径 $V \to Y$。则 c_{V} 属于任意图表 $D : X \to Y$。*3.* 在 $\mathcal{CD}(\mathcal{S})$中至多存在一个规范化网络图 $D : X \to Y$。*4.* 若 $X \subseteq V^{\mathrm{in}}$ ，则在 $\mathcal{Cart}(\mathcal{S})$中至多存在唯一态射 $D : X \to Y$。

证明. (1)：根据定义，任何规范化网络图都必须满足前两个条件，否则该图表将包含两条并非仅由复制映射连接的 V 导线，而最后一个条件是规范化所必需的。反之，假设这些条件成立。根据定义，X、Y 不含重复变量，因此任何输入和输出至多出现一次，且 D 由必要的部分构成。设 V 为该图表的一个输入。由于它不重复，且不能作为任何盒子的输出出现，所有 V 导线都必须由复制映射连接。假设 V 不是输入且出现在图表中。则根据签名的性质，V 必须作为某个 $\mathrm{c_V}$ 盒子的输出出现，且在 V 与 $\mathrm{c_V}$ 之间没有其他盒子，因为没有其他盒子具有输出类型 V，因此这必须可归约为一系列复制映射。此外，根据假设存在唯一这样的盒子 $\mathrm{c_V}$ ，因此这涵盖了 V 的所有实例。故 D 确实是一个网络图。

(2)：假设存在这样一条路径 $\mathrm{V} \to \mathrm{Y_i}$ ，其中 $\mathrm{Y_i} \in \mathrm{Y}$。我们通过对这类路径的最短长度 n 进行归纳法，证明 $\mathrm{c_V}$ 会出现在 D 中。

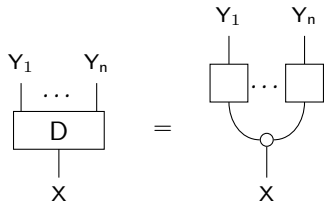
若 $n = 0$ ，则 $\mathrm{V} \in \mathrm{Y}$，且 $\mathrm{V} = \mathrm{Y_i}$。由于 $\mathrm{Y_i}$ 不是图表的输入，它必定是某个盒子的输出，因此 $\mathrm{c_{Y_i}}$ 必定出现在图表中。假设该结论对 n成立。那么，如果 V 有一条长度为 $(n+1)$ 的最短路径，它就有一个子节点 W，其最短路径长度为 n ，因此 $\mathrm{c_W}$ 会出现在 D 中。接着，D 中会出现一条 V 线；同理，由于 V 不是图表的输入，$\mathrm{c_V}$ 必定出现在 D 中。

(3)：设 $\mathrm{D}: \mathrm{X} \to \mathrm{Y}$ 是这样一个规范化网络图。由于 D 是网络图，对于它的每个输入 $\mathrm{X_i} \in \mathrm{X}$，$\mathrm{c_{X_i}}$ 都不可能出现在 D 中。因此，不失一般性地，我们可以假设每个 $\mathrm{X_i}$ 都是模型本身的输入。

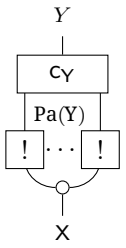
令 $G = G_{\mathbb{M}}$ 为对应于 $\mathbb{M}$ 的开放有向无环图（任意选择输出）。那么 D 也对应于一个开放有向无环图 H，其顶点是 G 中顶点的一个子集。此外，由于 D 是由 $\mathcal{S}$ 构建的，H 的边是 G 中边的一个子集；并且，如果 $V \in H \setminus$ 属于 (H)，那么 $\mathrm{Pa}^G(V) \subseteq H$ 中的所有边 $\mathrm{Pa}^G(\mathrm{V}) \to H$ 都在 D 中。因此，对于每个 $V \in H$，$\mathrm{Pa}^H(\mathrm{V}) = \mathrm{Pa}^G(\mathrm{V})$。

我们断言，H 恰好是 G 的全子有向无环图，由所有存在路径 $\mathrm{V} \to \mathrm{Y}$ 的顶点 V 给出，再加上 D 的任何其他输入，从而使得 H，因而 D 是唯一的。事实上，由于 D 是规范化的，若 $\mathrm{V} \in H \setminus$在$(H)$ 中，则在 H中存在路径 $\mathrm{V} \to \mathrm{Y}$，因而在 G中也存在这样的路径。反之，若 V 在 G 中有一条通向 Y 的路径，则由 (2)，c_{V} 出现在 D 中，因此 $\mathrm{V} \in H$，符合要求。

(4)：存在唯一的图表 $\mathrm{D} : \mathrm{X} \to ()$，由 $_X$给出。由于 $\mathbf{S}_{\mathsf{cart}}$中的每个态射都是确定性的，任何图表 $\mathrm{D}: \mathrm{X} \to (\mathrm{Y}_1$、…、$\mathrm{Y}_n)$ 都具有如下形式：



因此，只需考虑 $n = 1$的情形。我们通过对有向无环图使用归纳法来进行。若 Y 没有父节点，则不存在输出为 Y 的生成元，因此唯一的图表在交换的意义下由 $\mathrm{id}_Y \otimes$ 给出。现假设 $\mathrm{Pa}(\mathrm{Y})$ 非空，且该结果对 Y 的所有父节点成立。则 Y 不是该模型的输入，因而也不是该图表的输入，因此作为输出，它必须来自一个 c_{Y} 盒子。于是我们可以按如下方式对该图表进行因子分解，并且由归纳法可知这是其唯一形式。



□

命题 67。 设 π^S: $\mathbf{S}_H \to \mathbf{S}_L$ 是一个 *cd*-函子。记 $\pi(\mathrm{X}) := \pi^S(\mathrm{X})$ 在变量和类型 X上的映射。那么π^S 结构良好，当且仅当子集 $\pi(\mathrm{X})$ 两两不相交，且与 $\pi^S(\mathsf{V}_{\mathsf{H}}^{\mathsf{out}}) \subseteq \mathsf{V}_{\mathsf{L}}^{\mathsf{out}}$、$\pi^S(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}}) \subseteq \mathsf{V}_{\mathsf{L}}^{\mathsf{in}}$不相交，并且下图表可交换：

$$\begin{array}{ccc} \mathrm{Do}(\mathsf{V}_L) & \xleftarrow{\ \pi\ } & \mathrm{Do}(\mathsf{V}_H) \\ |-|_{\mathbb{L}} \downarrow & & \downarrow |-|_{\mathbb{H}} \\ \mathbf{S}_L & \xleftarrow{\ \pi^S\ } & \mathbf{S}_H \end{array} \tag{59}$$

当像命题 *58* 中那样定义 $\pi(\mathrm{Do}(\mathsf{S}): \mathsf{V}_{\mathsf{H}}^{\mathsf{in}} \otimes \mathsf{S} \to \mathsf{X}) := (\mathrm{Do}(\pi(\mathsf{S})): \pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}} \otimes \mathsf{S}) \to \pi(\mathsf{X}))$时。此外，这是唯一一个将查询映射为查询并使图表可交换的此类函子 π。

证明。 对于最后一个陈述，由于 $|-|_{\mathbb{L}}$ 和 $|-|_{\mathbb{H}}$ 在类型（变量）上是恒等映射，π 和 π^S 在对象上必须一致。因此，上述是从查询到查询唯一类型正确的函子映射，所以函子 π 若存在则唯一。

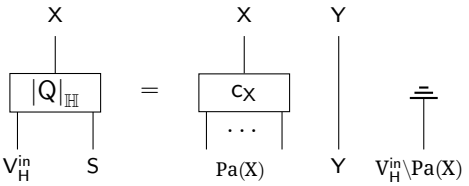
假设方形 (59) 可交换，我们声称每个 $\pi^S(\mathsf{c}_{\mathsf{X}})$ 都必须是规范化网络图。事实上，考虑查询 $\mathrm{Q} = (\mathrm{X} \mid \mathsf{V}_{\mathsf{H}}^{\mathsf{in}}, \mathrm{Do}(\mathrm{Y}))$，其中 $\mathrm{Y} := \mathrm{Pa}(\mathrm{X}) \setminus \mathsf{V}_{\mathsf{H}}^{\mathsf{in}}$。根据定义，这等于$|\mathrm{Q}|_{\mathbb{H}} = \mathsf{c}_{\mathrm{X}} \otimes \mathsf{V}_{\mathsf{H}}^{\mathsf{in}} \setminus \mathrm{Pa}(\mathrm{X})$。但于是 $|\pi(\mathrm{Q})|_{\mathbb{L}} = \pi^S(|\mathrm{Q}|_{\mathbb{H}}) = \pi^S(\mathsf{c}_{\mathrm{X}}) \otimes \pi(\mathsf{V}^{\mathsf{in}} \setminus \mathrm{Pa}(X))$，因此后者必定是规范化网络图，从而 $\pi^S(\mathsf{c}_{\mathrm{X}})$ 也必定如此，正如所需。

假设 π^S 满足上述条件。根据定义，结构良构性要求$(\pi(\mathrm{X}))_{X \in \mathsf{V}_H}$ 满足上述子集条件和不相交条件。根据假设，每个 $\pi^S(\mathsf{c}_{\mathsf{X}})$ 都是规范化网络图。此外，我们声称它包含非输入变量 $\alpha(\mathrm{X})$。确实，假设 c_{Y} 出现在

图表中。那么图表中必定存在一条从 $\mathbf{Y}$ 到输出 $\pi(\mathbf{X})$ 的路径。由于它是网络图，这条路径不能经过图表的输入，因此 $\mathbf{Y} \in \alpha(\mathbf{X})$。反过来，如果 $\mathbf{Y} \in \alpha(\mathbf{X})$，那么根据引理 66 (2)，$\mathbf{c_Y}$出现在 $\pi^S(\mathbf{c_X})$ 中。

现在我们证明，对于每种结构类型（笛卡尔、马尔可夫、cd），方形 (59) 可交换，当且仅当 π 在相应意义下也结构良好。

笛卡尔情形 Cart($\mathcal{S}$): 假设 π 不强。令 $\mathbf{V} \in \pi(\mathbf{Y}) \cap \alpha(\mathbf{X})$。那么 $\mathbf{Y} \notin \mathbf{Pa}(\mathbf{X})$。令 $\mathbf{S} := \mathbf{Pa}(\mathbf{X}) \cup \mathbf{Y} \backslash \mathbf{V}_{\mathbf{H}}^{\mathbf{in}}$。考虑查询 $\mathbf{Q} = (\mathbf{X}, \mathbf{Y} \mid \mathbf{V}_{\mathbf{H}}^{\mathbf{in}}, \mathbf{Do}(\mathbf{S}))$，即按 (60) 中的方式定义，但现在 $\mathbf{Y}$ 同时作为输出和输入。那么：



那么 $\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 包含两个互不相连的 $\mathbf{V}$ 实例，一个位于 $\pi^S(\mathbf{c_X})$ 内部，另一个位于 $\pi(\mathbf{id_Y})$ 内部。因此它不是网络图。

反过来，假设 π 强。考虑任意查询 $\mathbf{Q} = (\mathbf{X} \mid \mathbf{V}_{\mathbf{H}}^{\mathbf{in}}, \mathbf{Do}(\mathbf{S}))$。我们将证明，对于任意 $\mathbf{V} \in \pi(\mathbf{S})$，$\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 都不包含 $\mathbf{c_V}$。事实上，如果 $\mathbf{c_V}$ 属于该图，那么 $|\mathbf{Q}|_{\mathbb{H}}$ 中必定出现某个 $\mathbf{c_Y}$，且 $\mathbf{c_V}$ 属于 $\pi^S(\mathbf{c_Y})$，即 $\mathbf{V} \in \alpha(\mathbf{Y})$。但若 $\mathbf{c_Y}$ 以这种方式出现，则 $\mathbf{Y} \notin \pi(\mathbf{S})$。由于根据强性有 $\alpha(\mathbf{Y}) \cap \pi(\mathbf{S}) = \varnothing$，因此 $\mathbf{V} \notin \pi(\mathbf{S})$。由此可知，$\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 和 $|\pi(\mathbf{Q})|_{\mathbb{L}}$ 都是同一类型的（网络）图，且对于任意 $\mathbf{V} \in \pi(\mathbf{S})$ 均不包含 $\mathbf{c_V}$。因此，令 $\mathbb{M}' := \mathbf{Open_S}(\mathbb{L})$，我们可以将它们都视为 $\mathbf{S}_{\mathbb{M}'}$ 中的图，其中所有输入都是模型 $\mathbb{M}'$ 的输入。于是根据引理 66，它们在 $\mathbf{S}_{\mathbb{M}'}$ 中相等，因此在 $\mathbf{S}_L$ 中也相等。

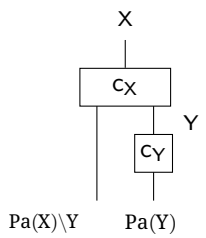
马尔可夫情形 Markov($\mathcal{S}$): 假设 π 不是超强的。则存在非输入变量 $\mathbf{X}$、$\mathbf{Y} \in \mathbf{V}_H$，且 $\mathbf{V} \in \alpha(\mathbf{X}) \cap \alpha(\mathbf{Y})$。于是机制 $\mathbf{c_V}$ 同时出现在 $\pi^S(\mathbf{c_X})$、$\pi^S(\mathbf{c_Y})$ 中，因此对于 $\mathbf{S}_H$ 中任何同时包含 $\mathbf{c_X}$、$\mathbf{c_Y}$的网络图 $\mathbf{D}$，$\pi^S(\mathbf{D})$ 包含 $\mathbf{c_V}$ 的两个实例，因而不是网络图。

反之，假设 π 是超强的。注意，在马尔可夫情形中，每个图表都会自动规范化。我们将证明，对于 $\mathbf{S}_H$ 中的任何网络图 $\mathbf{D}$，$\pi^S(\mathbf{D})$ 仍是 $\mathbf{S}_L$中的网络图。事实上，对于 $\mathbf{S}_H$中的任何此类图表 $\mathbf{D}$，每个 $c_{\mathbf{Y}}$ 至多出现在一个 $\pi^S(\mathbf{c_X})$中，因为由于 π是超强的，各 $\alpha(\mathbf{X})$ 彼此不相交，因而在 $\pi^S(\mathbf{D})$中至多出现一次。如果它没有通向输出的路径，则在马尔可夫情形中会自动从图表中移除。因此，由引理 66 (1)，$\pi^S(\mathbf{D})$ 是网络图。特别地，对于每个查询 $\mathbf{Q} \in \mathbf{Do}(\mathbf{V}_H)$，$\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 是网络图，于是由引理 66 (3)即可完成证明。

Cd情形 CD($\mathcal{S}$): 假设满性不成立，因此对于某个 $\mathbf{X}$，有 $\mathbf{Y} \in \mathbf{Pa}(\mathbf{X})$，其中 $\mathbf{Y}$ 不是输入；并且存在某个 $\mathbf{V} \in \pi(\mathbf{Y})$，但不存在从 $\mathbf{V}$ 到 $\mathbf{X}$ 的路径。考虑该查询

$$\mathbf{Q} = (\mathbf{X} \mid \mathbf{V}_{\mathbf{H}}^{\mathbf{in}}, \mathbf{Do}((\mathbf{Pa}(\mathbf{Y}) \backslash \mathbf{V}_{\mathbf{H}}^{\mathbf{in}}) \cup (\mathbf{Pa}(\mathbf{X}) \backslash \mathbf{Y})))$$

根据构造，$|\mathbf{Q}|_{\mathbb{H}}$ 是以下图表。



于是 $\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 包含 $\pi^{\mathbf{S}}(\mathbf{c_Y})$，而它又包含 $\mathbf{c_V}$。然而，由于不存在路径 $\mathbf{V} \to \pi(\mathbf{X})$，图表$\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 中不存在从 $\mathbf{c_V}$ 到输出的路径，因此这不是一个规范化网络图，所以不等于 $|\pi(\mathbf{Q})|_{\mathbb{L}}$。因此，满性是必要的。根据马尔可夫情形，额外的强性也是必要的，因为如果 π^S存在，它就限制为自由马尔可夫范畴之间的一个函子。

现在假设这两个条件都成立。对于 $\mathrm{Do}(\mathbf{V}_H)$ 中的任意查询 Q，马尔可夫情形的证明表明$\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 是一个网络图；接下来我们只需证明满性可确保它也是规范化的，然后根据引理 66 (3) 即可得证。

但需要注意的是，每个图表 $\pi^{\mathbf{S}}(\mathbf{c}_{\mathbf{X}})$ 都不涉及丢弃，因为它是规范化的，而满性保证图表的每个输入都有一条通往输出的路径。现在，对于任意查询 Q，$|\mathbf{Q}|_{\mathbb{H}}$ 都是规范化网络图，因此（除输入交换外）可分解为 $\mathbf{D}' \otimes$ ，其中 $\mathbf{D}'$ 是不涉及丢弃的网络图。于是 $\pi^S(|\mathbf{Q}|_{\mathbb{H}}) = \pi^S(\mathbf{D}') \otimes$ ，此时 $\pi^S(\mathbf{D}')$ 同样不涉及丢弃，因为每个$\pi^{\mathbf{S}}(\mathbf{c}_{\mathbf{X}})$ 都没有丢弃。因此，$\pi^S(|\mathbf{Q}|_{\mathbb{H}})$ 也经过规范化，符合要求。 □

现在，我们可以证明刻画机制级因果抽象的主要结果。

定理 *51*。 (3) $\Longrightarrow$ (1)：给定满足这些条件的 (π^S, τ)，根据命题 67，存在唯一的π: $\mathrm{Do}(\mathbf{V}_H) \to \mathrm{Do}(\mathbf{V}_L)$，使得 (π, π^S, τ) 构成机制级抽象，因而也是机制CCA。

(1) $\Longrightarrow$ (2)：根据假设 (π, π^S, τ)，(π, τ) 是一个构造性因果抽象。由于前者是机制级抽象，方形 (59) 可交换，因此 π^S 结构良好，进而对于每种结构类型，对齐 π 也结构良好。

(2) $\Longrightarrow$ (1)：设 (π, τ) 是一个构造性抽象，且 π 结构良好。则根据 π，对查询的作用方式为 $\mathrm{Do}(S) \mapsto \mathrm{Do}(\pi(S))$。

我们声称存在唯一一个函子 π^S，它结构良好，并且对于所有变量 X，都与 $\pi(\mathbf{X}) = \pi^S(\mathbf{X})$一致。

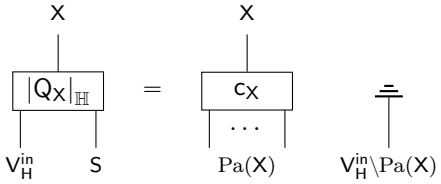
事实上，如果 π^S 结构良好，那么根据命题 67，对于查询映射 π: $\mathrm{Do}(S) \mapsto \mathrm{Do}(\pi(S))$，方形 (59) 可交换；这确实就是定理25所述构造性抽象中 π 的定义。因此该方形必定可交换。

现在，对于每个非输入变量 $\mathbf{X} \in \mathbf{V}^{\mathsf{in}^\perp}_{\mathsf{H}}$，我们必须定义 $\pi^S(\mathbf{c}_{\mathbf{X}})$ in (50)。定义 $\mathbf{S} := \mathrm{Pa}(\mathbf{X}) \backslash \mathbf{V}^{\mathsf{in}}_{\mathsf{H}}$，并定义查询

$$\mathbf{Q}_{\mathbf{X}} := (\mathbf{X} \mid \mathbf{V}^{\mathsf{in}}_{\mathsf{H}}, \mathrm{Do}(\mathbf{S})) \qquad \pi(\mathbf{Q}_{\mathbf{X}}) = (\pi(\mathbf{X}) \mid \mathbf{V}^{\mathsf{in}}_{\mathsf{L}}, \mathrm{Do}(\pi(\mathbf{S}))) \tag{60}$$

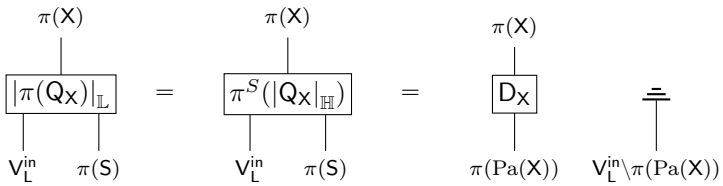
利用 $\mathbf{V}^{\mathsf{in}^\perp}_{\mathsf{H}} \subseteq \mathbf{V}^{\mathsf{out}}_{\mathsf{H}}$来定义前一个查询，并利用 $\pi(\mathbf{V}^{\mathsf{out}}_{\mathsf{H}}) \subseteq \mathbf{V}^{\mathsf{out}}_{\mathsf{L}}$定义后一个查询，因为 π 属于不相交变量对齐。

那么 $|\mathbf{Q}_{\mathbf{X}}|_{\mathbb{H}}$ 是一个规范化网络图，并且在 $\mathbf{S}_H$ 中我们有：



(61)

现在，由于 (59) 可交换，我们必须有：



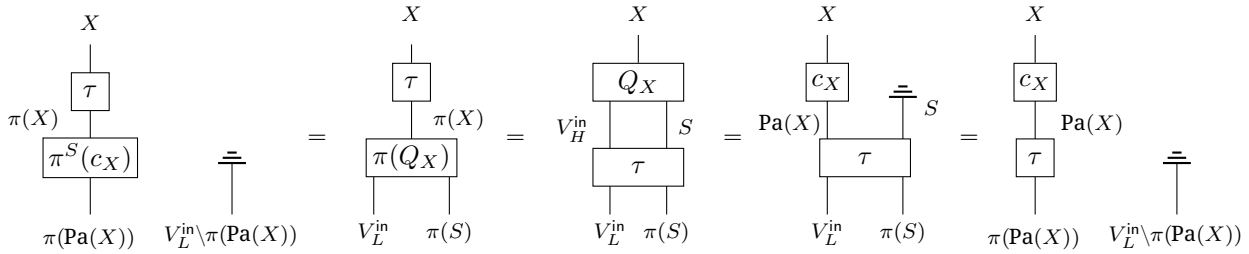
(62)

对于图表 $\mathbf{D}_{\mathbf{X}} := \pi^S(\mathbf{c}_{\mathbf{X}})$，最后一步使用了 (60)，并且 π^S 是一个 cd-函子，且 $\pi(\mathbf{V}^{\mathsf{in}}_{\mathsf{H}}) = \mathbf{V}^{\mathsf{in}}_{\mathsf{L}}$。现在我们声称，存在任何如右端所示的此类因子分解，等价于以下陈述：

$$\alpha(\mathbf{X}) \cap \mathbf{V}^{\mathsf{in}}_{\mathsf{L}} = \varnothing \tag{63}$$

事实上，根据定义，左端是一个规范化网络图，因此每个没有通向 X 的路径的输入变量都必须直接丢弃。于是，(63) 精确地表明其余部分包含于 $\pi(\mathrm{Pa}(\mathbf{X}))$ 中，即我们有这样的因子分解。因此，为了定义 π^S，(63) 必须对所有 $\mathbf{X} \notin \mathbf{V}^{\mathsf{in}}_{\mathsf{H}}$ 都成立；在这种情况下，$\pi^S(\mathbf{c}_{\mathbf{X}})$ 必须是 (62) 中唯一的图表 $\mathbf{D}_{\mathbf{X}}$。这使得 π^S 在存在时是唯一的。注意，若 π 是强的，则该性质成立，因此 π^S 可定义，因为 $\pi(\mathbf{V}^{\mathsf{in}}_{\mathsf{H}}) = \mathbf{V}^{\mathsf{in}}_{\mathsf{L}}$。因此，只要 π 是

结构良好的。最后，我们必须检验 τ 确实满足自然性 (51)。但在 $\mathbf{C}$ 中，我们有：



利用 (61)、(62) 以及 τ 关于查询的 cd-自然性。最后，根据我们对 $\mathbf{C}$ 的假设，将任意规范化状态作用于右侧丢弃态会将其抵消，从而得到 $\pi_X \circ \pi^S(c_X) = c_X \circ \pi$。因此， τ确实限制为所需的此类 cd-自然变换。□

引理 *52.* 根据命题 67，该条件显然是必要的，因为对于由 $\mathsf{Q} = (\mathsf{V}_\mathsf{H}^\mathsf{out} \mid \mathsf{V}_\mathsf{H}^\mathsf{in}、\mathsf{Do}())$ 定义的查询，有 $\mathsf{D}_\mathbb{H} = |\mathsf{Q}|_\mathbb{H}$；因此，若 (59) 可交换，则 $\pi^S(\mathsf{D}_\mathbb{H}) = |\pi(\mathsf{Q}|_\mathbb{L}$，从而构成一个规范化网络图。根据定义， π^S 结构良好意味着它满足这些条件。事实上，注意到由于 $\pi^S(\mathsf{V}_\mathsf{H}^\mathsf{in}) = \mathsf{V}_\mathsf{L}^\mathsf{in}$ 且所有 $\pi^S(\mathsf{X})$ 两两不相交，因此必有 $\pi^S(\mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{H}) \subseteq \mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{L}$。

反过来，我们现在证明这些条件也是充分的。假设这些条件成立。由于 $\pi^S(\mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{H}) \subseteq \mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{L}$，因此对于 $\mathsf{X} \in \mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{H}$，各个 $\pi^S(\mathsf{X})$ 是两两不相交的子集。由于 $\pi^S(\mathsf{V}_\mathsf{H}^\mathsf{in}) = \mathsf{V}_\mathsf{L}^\mathsf{in}$，对于 $\mathsf{X} \in \mathsf{V}_\mathsf{H}^\mathsf{in}$ 也同样成立。由于 $\mathsf{V}^{\mathsf{in}^\perp}{}_\mathsf{L}$ 与 $\mathsf{V}_\mathsf{L}^\mathsf{in}$ 不相交，因此各个 $\pi^S(\mathsf{X})$ 也就两两不相交。

我们需要证明每个 $\pi^S(\mathsf{c_X})$ 都是网络图。由于 $\pi(\mathsf{X})$ 彼此不相交，该图表没有重复的输入或输出。现在假设有两个 $\mathsf{c_V}$ 的实例属于 $\pi^S(\mathsf{c_X})$（规范化后）。由于它们都存在一条通往 $\pi^S(\mathsf{c_X})$ 输出的路径，因此它们也都存在一条通往 $\pi^S(\mathsf{D}_\mathbb{H})$ 输出的路径，因而无法被丢弃，这就意味着后者不是网络图，矛盾。接下来假设 $\mathsf{V} \in \pi(\mathsf{Y})$，其中 $\mathsf{Y} \in \mathsf{Pa}(\mathsf{X})$。那么 Y 不能是输入（否则 V 就会是输入，因此 $\mathsf{c_V}$ 不属于签名）。根据引理 66 (2)，我们有 $\mathsf{c_V} \in \pi^S(\mathsf{c_Y})$。于是 $\mathsf{c_V}$ 同时出现在 $\pi^S(\mathsf{c_X})$ 和 $\pi^S(\mathsf{c_Y})$ 中（这是由它们的输入和输出类型决定的），并分别存在通往两者输出 $\pi(X)$ 和 $\pi(Y)$ 的路径。由于这些是 $\pi^S(\mathsf{D}_\mathbb{H})$ 的输出，这两个 $\mathsf{c_V}$ 实例都无法从图表 $\pi^S(\mathsf{D}_\mathbb{H})$ 中移除（即不会被丢弃），因此该图表不可能是网络图，矛盾。因此，根据引理 66 (1)，每个盒子 $\pi^S(\mathsf{c_X})$ 确实都是网络图。

反之，假设 π 不是超强的。则存在非输入变量 $\mathsf{X}、\mathsf{Y} \in \mathsf{V}_H$ 以及某个 $\mathsf{V} \in \alpha(\mathsf{X}) \cap \alpha(\mathsf{Y})$。根据引理 66 (2)，$\mathsf{c_V}$ 同时出现在 $\pi^S(\mathsf{c_X})$ 和 $\pi^S(\mathsf{c_Y})$ 中，并且正如上文所述，这使得$\pi^S(\mathsf{c_X})$ 不是网络图，从而产生矛盾。□

B 代数结构的保持

将干预组织成诸如 $\mathcal{I}^{\mathrm{io}}(\mathbb{V})$、$\mathrm{Do}(\mathbb{V})$ 和 $\mathrm{Inc}(\mathbb{V})$ 之类的查询签名，对于形式化因果抽象一直很有用。然而，（由它们生成的自由范畴）往往几乎没有有趣的组合结构，除查询本身之外，唯一的新复合项只是查询的乘积。

然而，干预具有代数结构还有第二层含义：我们可以依次相继地应用它们。由于这一点在因果抽象中发挥了作用，自然要问我们能否在本框架中刻画它。对于一般模型和查询，一种有用的做法如下。

定义 *68*。我们称查询签名 $\mathcal{Q}$ 具有幺半群结构，是指存在一个具有相同类型的查询签名 $\mathcal{T}$ 以及一个幺半群 $(M_\mathcal{Q}、\bullet)$，使得 $\mathcal{Q}$ 具有查询 $\mathcal{T}_{mor} \times M_\mathcal{Q}$。我们称 $\mathcal{Q}$ 构成一个幺半群，是指 $\mathcal{T}$ 具有两个类型 $\mathit{in}、\mathit{out}$，并且恰好有一个查询 $q\colon \mathit{in} \to \mathit{out}$。对于两个具有幺半群结构的查询签名 $\mathcal{Q}、\mathcal{Q}'$，我们称映射 $F\colon \mathcal{Q} \to \mathcal{Q}'$ 是一个同态，是指它分解为 $F((Q, m)) = (F_Q(Q), F_M(m))$，其中 $F_M\colon M_\mathcal{Q} \to M_{\mathcal{Q}'}$ 是一个幺半群同态。

让我们首先考察一些具有幺半群结构的查询的关键例子。

示例 69。 设 $\mathbb{V}$ 为一个具体变量集合。

(1) 查询签名 $Do(V)$ 具有幺半群结构 $io(V) \times \mathbb{P}(V^{in^\perp})$，其中 $io(V)$ 表示 V 上的输入输出查询，而 $\mathbb{P}(V^{in^\perp})$ 表示 V 的子集在 $S \bullet T := S \cup T$ 运算下构成的幺半群。在这种情况下，我们实际上可以通过 $(O \mid V^{in}、Do(S)) \bullet (O \mid V^{in}、Do(T)) = (O \mid V^{in}$，将幺半群乘法扩展为 $Do(V)$ 上的部分运算 $Do(S \cup T))$；当两个查询的输出子集一致时，该运算有定义。

(2) 查询签名 $\mathrm{Do}(\mathbb{V})$ 通过平凡的"空"干预 $\mathbf{1}$ 和复合$\mathrm{Do}(X = x) \bullet \mathrm{Do}(Y = y) := \mathrm{Do}(X = x, Y \backslash X = y|_{Y \backslash X})$ 构成一个幺半群。

(3) 更一般地，对于任意干预集 $\mathcal{I}$，如果查询签名 $\mathcal{I}^{\mathrm{io}}$ 对复合封闭，则其构成一个幺半群：对于两个干预 σ, σ'，它们对应的"新"机制集合分别为 $\{c_X\}_{X \in A_\sigma}$ 和$\{c'_Y\}_{Y \in A_{\sigma'}}$，复合干预 $\sigma \circ \sigma'$ 定义为 $\{c_X\}_{X \in A_\sigma} \cup \{c'_Y\}_{Y \in A_{\sigma'} \backslash A_\sigma}$；单位元为 $A_\sigma = \varnothing$。

第5节中的许多因果抽象概念确实保留了相应干预集合中存在的这种代数结构。以下内容很容易验证，因此我们省略证明。

命题 70。 根据例 *69* 中给出的幺半群结构，有：

(a) 对于任意构造性因果抽象，查询映射 $\pi \colon \mathrm{Do}(\mathsf{V_H}) \to \mathrm{Do}(\mathsf{V_L})$ 都是同态；它所定义的向上抽象，其查询映射 $\omega \colon \mathrm{Do}(\mathbb{V}_L) \to \mathrm{Do}(\mathbb{V}_H)$ 也是同态。[49]

(b) 任意同构构造　性抽象定义了一种整体向上抽象，其映射 ω 是同态　　$m.$

同态与保序。 给定具有幺半群结构的查询签名之间的一个精确变换（定义 21），也就是说，这些签名在干预的复合下封闭，查询映射 ω 可能是同态，也可能不是。在文献中，实际上要求该映射在偏序 $\leqslant$ 下满足保序 性质：对于 $\sigma \leqslant \sigma'$，只要 $\sigma \circ \sigma' = \sigma'$ [RWB+17，就有 +23]，BH19，GIZ[50] 。GIZ [的研究通过考察干预的代数结构，将干预之间的交换关系也视为重要因素。成为同态意味着同时保持交换关系和序关系：对于后者，只要 $\sigma \circ \sigma' = \sigma'$，就有 $\omega(\sigma') = \omega(\sigma \circ \sigma') = \omega(\sigma) \circ \omega(\sigma')$。因此，或许直接要求 ω 是同态更为自然。这也可能为进一步强化强因果抽象 （定义 23）提供一个自然条件；本研究中所有后者的示例都满足这一条件。

C 与文献的比较

备注 71 (精确变换**)**。 定义 *21* 中给出的精确变换概念与文献中的概念存在以下差异。

1. 代数结构：如 *B* 节所详细讨论的，文献假设干预上的映射 ω 是保序映射，而我们不作此假设。

2. 满射性：在 *[RWB+17、BH19]* 中（不同于 *[GIZ+23]*），精确变换的映射 τ 不要求是满射。然而，*[BH19]* 中证明了，若要称为因果抽象τ，就应要求其为满射。事实上，我们认为这是任何抽象关系（无论是否因果）的必要要素，因此 τ 对第 *4* 节中的两种基本抽象都必然是满态射。在这方面，我们的精确变换概念更接近于 *[BH19]* 中的 τ-抽象。

3. 因果模型的类型：除了反事实抽象之外，我们的定义都不对因果模型的种类作出假设——只要求不存在被视为根本隐藏的变量：变量要么是输入，要么不是；虽然 $\mathsf{V}^{\mathsf{in}^\perp} = V \backslash \mathsf{V}^{\mathsf{in}}$ 不一定全都是输出，但原则上可以对其中任意变量进行干预。[51] 鉴于任何因果抽象的概念

[49]这里的 π 是签名之间的映射；根据定义 15 中的约束，它由构造性因果抽象的函子 π 确定。并且，称相应向上抽象的映射 ω: $\mathrm{Do}(\mathbb{V}_L) \to \mathrm{Do}(\mathbb{V}_H)$ 为同态，意指它在 ω 的定义域上是同态；按构造，该定义域确实对复合封闭。[50]在 [RWB+17] 和 [BH19], 中，由于它们不考虑比 Do-干预更一般的干预，所要求保持的偏序是 $\mathrm{Do}(S = s) \leqslant \mathrm{Do}(S' = s')$ 当且仅当 $S \subseteq S'$ *and* $s'|_S = s$。显然，相对于该偏序的保序性蕴含了对上述偏序的保持。[51]因此， [BH19], 中的一个关键观察——即允许其意义上的"概率模型"可能导致对根节点上的概率分布进行微调所造成的"不当抽象"——在此并不适用。

都只有其所适用的（干预）查询集合所赋予的“强度”，从一开始就限制模型类型似乎并非必要。

4. 干预的种类：与 *[RWB+17* 不同，*BH19]* 在定义 *21* 中并未将 $\mathcal{I}_L, \mathcal{I}_H$ 限定为 *Do*-干预。（在这方面，我们的观点更接近 *[GIZ+23]* 中的概念。）

注 72（强因果抽象）。虽然我们为任意干预定义了精确变换，但 *Do*-干预在概念上意义最为重大，因为只要能够获取所有这些干预，就可以识别出完整的因果模型。因此，模型 $\mathbb{L}$ 与 $\mathbb{H}$ 之间因果抽象的“标准”实例，是指 $\mathcal{I}_H$ 包含所有 *Do*-干预的情形——*Beckers and Halpern* 随后将他们的 τ-抽象概念称为“强”*[BH19]*。我们的强因果抽象概念（定义 *23*）与此一脉相承，但也有几处不同。

1. 与精确变换的情形（见上文）一样，我们不对模型的类型作任何限制，也不作保序假设。*2.* 更重要的是，虽然 $\mathcal{I}_H$ 是所有 *Do*-干预的集合，但 $\mathbb{L}$ 原则上不受约束，因此，关于查询的映射 ω 是给定数据的一部分，而不像 *[BH19]* 中那样由 τ 导出。

其动机源于多方面的考量。首先，基于神经网络的现代人工智能拓宽了因果抽象的范围，因此我们应当预期，可解释变量的低层表示会呈现“分布式”特征。其次，从理论角度来看，我们希望在一般干预的背景下理解因果抽象，而因果模型框架本身就涵盖了一般干预——这与机器学习无关。事实上，*Beckers and Halpern* 在 *[BH19],* 中提出了一个猜想，其基本含义是，“强性”（所有高层*Do*-干预）意味着“构造性”，即存在一种如不相交变量对齐所示的划分。关键在于，这一直觉从一开始就依赖于将低层干预也限制为 *Do*-干预。我们不难验证，在我们对强因果抽象的设定中，这种限制等价于找到$\mathbf{V}_L$的一个划分。虽然鉴于双方设定中提及的其他差异，这尚未回答他们的猜想，但它表明，我们的概念为进一步理解因果抽象提供了一个有用的视角。

C.1 与 *Englbergerand Dhami* 的比较

与此密切相关的是 Englberger and Dhami [ED25], 的工作，他们基于马尔可夫范畴和自然变换，对因果抽象进行了类似的形式化。然而，尽管他们对因果抽象的范畴定义旨在捕捉构造性因果抽象（如我们的定理25所示），实际上（经过修正后），该定义更具体地捕捉了机制级构造性抽象。

更具体地说，他们将因果抽象定义为一个函子 $\iota = \pi\colon \mathbf{S}_H \to \mathbf{S}_L$ 以及一个确定性自然变换 $\tau\colon [\![-]\!]_{\mathbb{L}} \circ \iota \implies [\![-]\!]_{\mathbb{H}}$，其中 $\mathbf{S}_H$、$\mathbf{S}_L$是由因果签名诱导的自由马尔可夫范畴，并且满足 ι 将 $\mathbf{restr}(\mathbf{S}_H)$ 中的态射映射到 $\mathbf{restr}(\mathbf{S}_L)$中的态射。这里，$\mathbf{restr}(\mathbf{S}_L)$ 被定义为 $\mathbf{S}_L$ 的子范畴，即“限制到每个生成盒至多出现一次的态射”后得到的子范畴。在他们的命题 2.1中，他们声称证明了对于所有 $\mathbf{X}$、$\mathbf{Y} \subseteq \mathbf{V}_\mathbf{L}$，存在唯一的此类态射 $\mathbf{X} \to \mathbf{Y}$。

然而，遗憾的是，情况并非如此。例如，考虑一个具有有向无环图 $\mathbf{X} \to \mathbf{Y}$ 以及互不相同的态射 $\mathbf{c}_\mathbf{Y} \neq \mathbf{c}_\mathbf{Y} \circ \mathbf{c}_\mathbf{X} \circ \doteq_X\colon \mathbf{X} \to \mathbf{Y}$ 的因果模型。人们可以转而将 $\mathbf{restr}(\mathbf{S}_L)$ 定义为由（规范化的）网络图构成，但它们形成的是一个签名而非一个范畴，从而恰好成为我们关于抽象 Do-干预的签名 $\mathbf{Do}(\mathbf{V}_L)$。于是，命题 2.1 对应于引理 66 (3) 的唯一性结果。

对这两个模型都做此替换后，他们对因果抽象的定义便恰好是针对查询集 $\mathbf{Do}(\mathbf{V}_L)$、$\mathbf{Do}(\mathbf{V}_H)$ 的组件级抽象（定义 46），即机制级构造性抽象（6.2）（除了不要求 τ 为满态射之外）。他们的定理 3.1 声称这些与一种形式的构造性因果抽象之间存在等价。

然而，由于上述细微之处，并非每个构造性抽象都属于他们所说的形式。事实上，根据我们的刻画定理 51，这一定义所刻画的是额外简单的构造性抽象。[52]

[52]另请注意，引理 52 基于自由马尔可夫范畴给出了机制层面抽象的进一步刻画；事实上，我们只需验证 $\pi^S(\mathbf{D}_\mathbb{H})$ 仍是一个网络图，而不必对 $\mathbf{Do}(\mathbf{V}_H)$ 的所有图表施加条件。

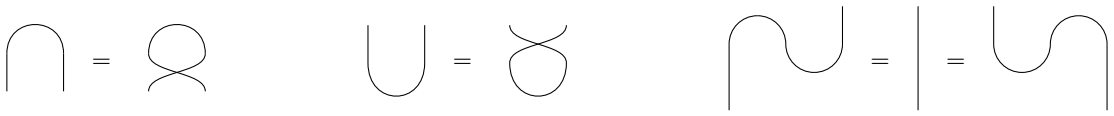
例 *73*。一个不符合 *[ED25]* 中因果抽象定义的构造性抽象示例，由例 *54* 中的模型 $\mathbb{L},\mathbb{H}$ 给出，这里令 $W = I$，此时 $a = c \circ e\colon I \to X$ 和 $b = d \circ e\colon I \to Y$，其中 e 是 Z 的一个尖锐状态。存在一个具有 $\tau = id$ 的构造性抽象，由 $\pi(X) = X$ 和 $\pi(Y) = Y$ 给出。若将其扩展为函子 $\pi^S = \iota\colon S_H \to S_L$，则必须定义为 $\pi^S(a) = c \circ e$ 和 $\pi^S(b) = d \circ e$。但这样一来，作为图表，我们就会有 $\pi^S(D_{\mathbb{H}}^{X\,Y}) \neq D_{\mathbb{L}}^{X\,Y}$，正如式 *(53)* 中那样。

注释 74. 正如我们在第 *8* 节所指出的，人们可能会想知道，许多或大多数具有实际意义的构造性抽象，最终是否可能确实都是机制层面的，也就是说，它们是额外简单的，并且确实符合 *[ED25]* 的（修正版）定义；我们将这项研究留待未来工作。

D 并行机制通道与一致性条件图

第5.5节引入了并行机制信道 $\mathcal{F}_{\mathbb{M}}$ 的概念，该信道由因果模型 $\mathbb{M}^{\rightarrow}$ 所诱导（反之亦然）。这里我们将进一步阐明 $\mathcal{F}_{\mathbb{M}}$ 与 $\mathbb{M}$ 之间的关系，以便表达分布式抽象的一致性条件。

为此，下面的概念将会很有用。回想一下，当每个对象 X 都配有一个对偶 $X^\star$，以及定义在 $X \otimes X^\star$ 和 $X^\star \otimes X$ 上的单位与余单位，并满足若干公理 [Sel10] 时，对称幺半范畴 $\mathbf{C}$ 就称为 紧闭 。为求简便，并且考虑到这适用于我们的主要示例，我们可以取 $X^\star = X$，此时定义可简化为：每个对象 X 都配有一对特殊状态和效应，分别定义在 $X \otimes X$、[53] 上，并分别表示为杯和帽，满足以下条件：



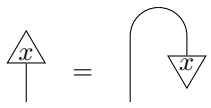
(64)

如果$\mathbf{C}$ 也是一个 cd-范畴，我们还要求：



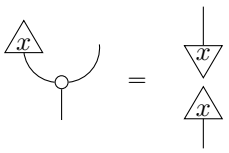
(65)

对于任意尖锐状态 x，都有一个对应的确定性效应，记作 $x^\dagger$，并描绘为“将 x上下翻转”：



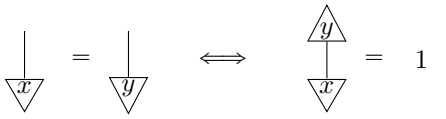
(66)

由 (65) (a) 可知，任意归一化尖锐状态 x 都满足 $x^\dagger \circ x = 1$；由 (b) 可知：



(67)

将 (67) 与 y 复合可得，对于所有尖锐状态 x、y，都有：

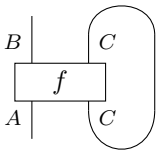


(68)

[53] 对象 X 上的 *effect* 是一个态射 $X \to I$，其 I 为单位对象。

我们的主要示例是，马尔可夫范畴 **FStoch** 嵌入在更广泛的紧闭 cd-范畴$\mathbf{Mat}_{\mathbb{R}^+}$ 中，其中对象仍然是有限集，而态射 $M\colon\ X \to Y$ 现在是任意的（不一定是随机的）正矩阵 $M(y \mid x)_{x\in X,\ y\in Y}$。这里的杯和帽都只是 $\delta_{X,\ X}$，即“完全相关”。

最后，我们回顾一下，紧闭性还会诱导出一个带迹范畴 [Sel10], 即偏迹概念的范畴化推广。对于每个对象 C 相应的杯和帽为任意对象 A、B定义映射 $\mathrm{Tr}_C\colon\ \mathbf{C}(A\otimes C,\ B\otimes C) \to \mathbf{C}(A,\ B)$，其中 $\mathrm{Tr}_C(f)$ 由下式给出

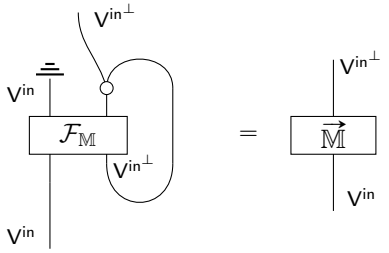


(69)

将所有这些部分结合起来，就得到了函数不动点与迹之间的关系；在我们的设定中，这体现为 $\mathcal{F}_{\mathbb{M}}$ 与 $\mathbb{M}$ 之间如下所示的关系。

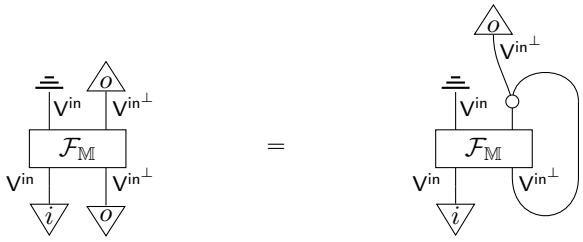
如果对于任意 $f\circ x = g\circ x$，对所有归一化尖锐态x，都有 $f = g$，则称 **C** 具有足够多的归一化尖锐态。

引理 75。 设**C** 是满足上述性质且具有足够多归一化尖锐态的紧闭 *cd*-范畴。则对于在**C** 中、关于变量 $\mathbb{V}$ 的任意因果模型 $\mathbb{M}$，其中$\mathsf{V}^{\mathsf{out}} = \mathsf{V}^{\mathsf{in}^\perp}$，都有：



(70)

证明。 设 i、o分别是 V中的、V中的尖锐状态$^\perp$。根据引理 40，我们有 $\vec{\mathbb{M}}\circ i = o$当且仅当 $\mathcal{F}_{\mathbb{M}}\circ(i\otimes o) = (i\otimes o)$。由于 $\mathcal{F}_{\mathbb{M}}$ 是 V^{in} 上的恒等映射，这等价于满足$\mathsf{v}^{\mathsf{in}}\circ\mathcal{F}_{\mathbb{M}}\circ(i\otimes o) = o$。根据（68），这又等价于下方左端等于 1，而下方的等式由（67）推出。

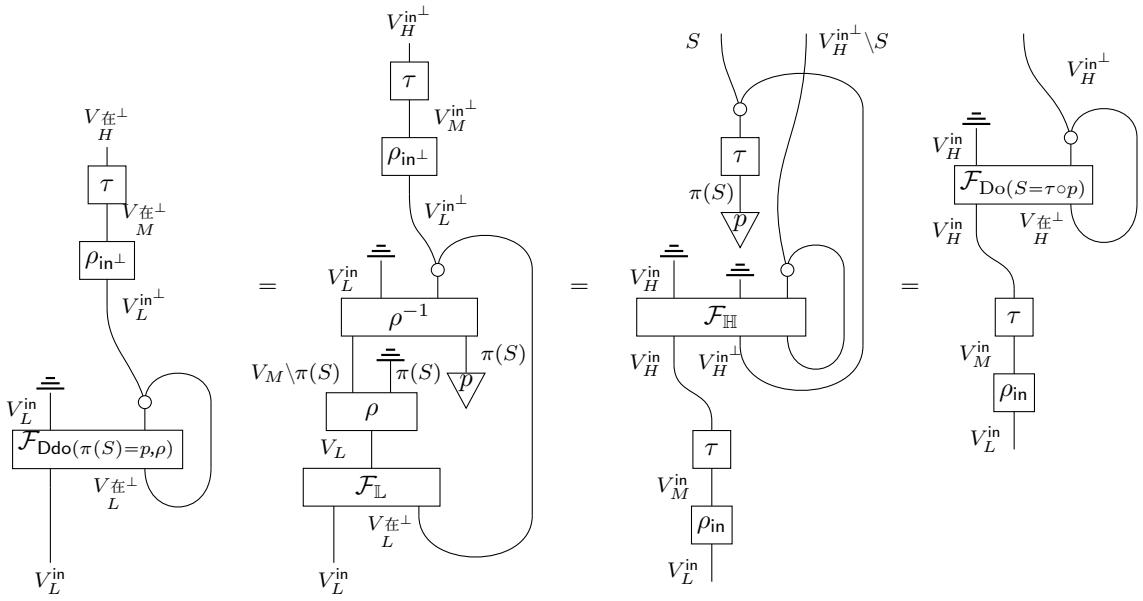


因此，上述右端等价于 1。再次应用 (68)，这等价于 (70) 的左端与 i 复合，得到 o。至此证明完成，因为这对任意状态 i都成立，而且**C** 具有足够多的状态。

□

我们现在可以写出第 5.5.2 节中同构构造性抽象的一致性条件 (44)，并考-

ering该 低层级分布式 Do-干预（来自式 (45)），采用如下更直接的形式 :



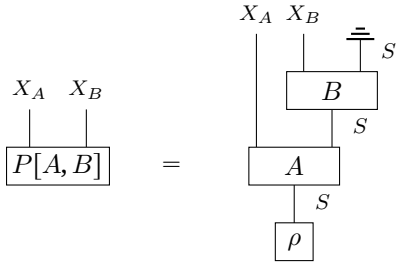
一致性条件 同构交换抽象的条件可以用类似的方式写成 ay.

备注 76. 在所谓的分裂节点因果模型 形式体系中，经典因果模型有一种等价的表示方式，其中因果模型的数据定义了一个高阶映射 *[BLO19, LT23]*。本文所称的模型 $\mathbb{M}$ 的并行机制过程 $\mathcal{F}_{\mathbb{M}}$，除了一些记账上的差异外，本质上与一种常见的分裂节点模型高阶映射表示方式相同。以如上方式使用迹$\overrightarrow{\ }$结构，在 $\mathcal{F}_{\mathbb{M}}$ 与模型的输入输出过程 $\mathbb{M}$ 之间转换，是在分裂节点模型形式体系中重新阐述第 *5* 节因果抽象概念的关键。鉴于分裂节点模型也是精确定义经典因果模型如何成为量子因果模型特例的方式 *[BLO19]*，“迹结构技巧”对于研究量子因果抽象概念也至关重要（参见第 *7* 节及即将发表的工作）。

E 量子-经典约化

以下示例说明了如何将量子模型 $\mathbb{L}$ 抽象为 **量子-经典**中的经典模型 $\mathbb{H}$，这可以看作是将量子模型约化为经典模型的一种“约化”。

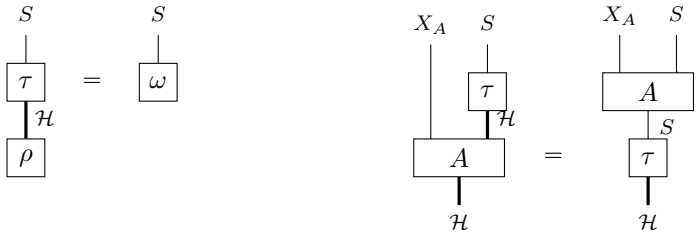
示例 77 （量子决策模型）。 *[TO26]* 考虑这样一项实验：人类受试者做出决定 A 和 B，结果为经典结果 X_A，X_B；例如，回答答案集为 $X_A = X_B =$ {是、否}的问题。这会产生分布 $P[A,\ B]$，对应于“A 然后 B”的答案序列，以及分布 $P[B,\ A]$，对应于“B 然后 A”。一个决策模型指定一个处于初始状态 ρ 的对象 S，以及每项决策对应的通道 A，B，使得在满足以下条件时，我们能够恢复这些分布：[54]



[54]因此，它是一个组合模型，其签名 $\mathcal{S}$ 由 S、A、B, ρ 组成，分别建模为 (S, A, B, ρ)，查询 $\mathcal{Q}$ 由 P[A, B]、P[B, A] 给出，建模方式如上。

以及类似的“B 然后 A”方程。量子决策模型，特别是**量子-经典**模型（其中 $[\![S]\!] = \mathcal{H}$），被提出用来解释各种心理效应 *[TO26]*。

给定这样的模型 $\mathbb{L}$，向由有限集 S、分布 ω 和通道 A, B给出的经典决策模型 $\mathbb{H}$进行抽象，会意味着什么？向下抽象 $\mathbb{L} \to \mathbb{H}$，其中 $\pi = \mathrm{id}_{\mathbf{Q}}$、$\tau_{X_A} = \mathrm{id}$ 和$\tau_{X_B} = \mathrm{id}$，将表明两个模型在 $P[A, B]$和 $P[B, A]$上是一致的，也就是说，实际上存在一个经典模型 $\mathbb{H}$。更强一些，带有 $\pi^S = \mathrm{id}_{\mathbf{S}}$ 的组件级抽象将给出一个通道 $\tau\colon \mathcal{H} \to S$，使得：



对于 B，情况也类似。最后，更进一步，还可以提供一个带有$\tau \circ \tau^{-1} = \mathrm{id}_S$的“制备”信道 $\tau^{-1}\colon S \to \mathcal{H}$，这样每个量子信道 A都会化简为一个“测量”信道 τ、经典信道 A和重新制备过程 τ^{-1}。例如，若将 τ，τ^{-1}作为正交归一基的测量信道和制备信道，则 A 和 B 在该基下均为对角形式。

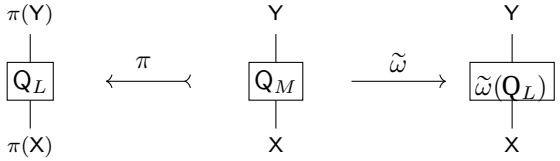
F 抽象的进一步形式化

F.1 向上抽象的形式化

现在，我们通过将向上抽象与自然变换联系起来，明确图表（16）成立的含义。

命题 78. 向上抽象 $(\pi,\ \tau,\ \omega)$ 等价于存在 $(\mathbb{M},\ \mathcal{Q}_M)$ 以及两个向下抽象：

$$(\mathbb{L}, \mathcal{Q}_L) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{M}, \mathcal{Q}_M) \underset{\widetilde{\omega}}{\overset{\mathrm{id}}{\rightleftarrows}} (\mathbb{H}, \mathcal{Q}_H)$$



(71)

其中，$(\mathbb{M},\ \mathcal{Q}_M)$ 是两个向下抽象中的“高层模型”，π 在查询上是单射，而 $\widetilde{\omega}$在类型上是恒等映射、在查询上是满射。与 $\omega\colon\ \mathcal{Q}_L \twoheadrightarrow \mathcal{Q}_H$ 的对应关系，通过$\omega(\mathsf{Q}_L) = \widetilde{\omega}(\mathsf{Q}_M)$ 在 $\mathcal{Q}_H$ 中实现，适用于任意满足 Q_L 的$\mathsf{Q}_L = \pi(\mathsf{Q}_M)$。[55]

上文中，每个查询 Q_M in $\mathcal{Q}_M$ 都具有来自 $\mathcal{Q}_H$的类型 X、Y，但也可以将其视为查询 $\mathsf{Q}_L := \pi(\mathsf{Q}_M)\colon \pi(\mathsf{X}) \to \pi(\mathsf{Y})$ in $\mathcal{Q}_L$，即一个在其上定义了 ω、且像为 $\omega(\mathsf{Q}_L) := \widetilde{\omega}(\mathsf{Q}_M)$ in $\mathcal{Q}_H$的查询。

证明。 假设我们已得到如上所述的向下抽象。由于 $\mathcal{Q}_H$ 中的类型与 $\mathcal{Q}_M$中的类型相同，我们也可以将 (π, τ) 视为 $\mathbb{L}$与 $\mathbb{H}$之间的类型对齐。现在，我们按照 (71) 中的说明，将 $\widetilde{\omega}$扩展到$\mathcal{Q}_L$。对于 $\mathsf{Q} \in \mathcal{Q}_L$，当且仅当存在某个 $\mathsf{Q}' \in \mathcal{Q}_M$使得 $\mathsf{Q} = \pi(\mathsf{Q}')$ 时，才将 $\omega(\mathsf{Q})$ 定义为有定义。在这种情况下，定义 $\omega(\mathsf{Q}) := \widetilde{\omega}(\mathsf{Q}')$。于是，由于 $\widetilde{\omega}$ 是从 $\mathcal{Q}_M$到上的满射，ω 也是从 $\mathcal{Q}_L$到上的满射；并且可以验证，每当它将一个查询映射到类型为 $\mathsf{X} \to \mathsf{Y}$ 的查询时，该查询的类型为 $\pi(\mathsf{X}) \to \pi(\mathsf{Y})$。

还需要检查一致性是否成立。假设对于 $\mathsf{Q} \in \mathcal{Q}_L$，$\mathsf{Q}_H = \omega(\mathsf{Q})$。那么对于 $\mathsf{Q}' \in \mathcal{Q}_M$，$\mathsf{Q} = \pi(\mathsf{Q}')$。因此，一致性成立，因为：

$$Q_H \circ \tau = \widetilde{\omega}(Q') \circ \tau = Q' \circ \tau = \tau \circ \pi(Q') = \tau \circ Q$$

[55]注意，来自向下抽象的函子 π 与向上抽象的 π 相一致，后者只是一个类型映射。

在第二个等式中，我们用到了 (τ, id) 是一个严格向下抽象。在第三个等式中，我们用到了 (π, τ) 是一个向下抽象。

反过来，给定任意向上抽象，定义 $\mathcal{Q}_M$ 使其具有与 $\mathcal{Q}_H$ 相同的类型，并且对每个 ω 有定义的查询 $\mathrm{Q}: \pi(\mathrm{X}) \to \pi(\mathrm{Y})$，定义一个查询 $\mathrm{Q}^*: \mathrm{X} \to \mathrm{Y}$；并对所有类型 X，令 $[\![\mathrm{Q}^*]\!]_{\mathbb{M}} := [\![\omega(\mathrm{Q})]\!]_{\mathbb{H}}$，且 $[\![\mathrm{X}]\!]_{\mathbb{M}} := [\![\mathrm{X}]\!]_{\mathbb{H}}$。定义 $\pi(\mathrm{Q}^*) := \mathrm{Q}$，并令 $\tilde{\omega}$ 由 $\tilde{\omega}(\mathrm{Q}^*) := \omega(\mathrm{Q})$ 定义，且在类型上为恒等映射。根据定义，π 在类型上是单射，而 $\tilde{\omega}$ 是满射。$[\![\mathrm{Q}^*]\!]_{\mathbb{M}} := [\![\omega(\mathrm{Q})]\!]_{\mathbb{H}}$ 这一事实保证 $(\tilde{\omega}, \mathrm{id})$ 构成严格向下抽象。最后，要使 (π, τ) 构成向下抽象，我们必须检查一致性；由于始终有 $[\![\mathrm{Q}^*]\!]_{\mathbb{M}} \circ \tau = [\![\omega(\mathrm{Q})]\!]_{\mathbb{H}} \circ \tau = \tau \circ \pi(Q)$，一致性成立。 □

F.2 抽象的刻画

我们可以用更范畴化的方式刻画交换抽象和反事实抽象，如下所示。

引理79。 对于结构范畴为**S**的任意 $\mathbb{V}$ 上的因果模型，将每个查询映射到其图表因子的函子 $|-|: \mathsf{Inc}(\mathsf{V}) \to \mathbf{S}$ 可经由函子 $|-| : \mathsf{Inc}(\mathsf{V}) \to \mathrm{Do}(\mathsf{V})$，对 *Do* 查询分解为 $|-|: \mathrm{Do}(\mathsf{V}) \to \mathbf{S}$。[56] 于是，交换抽象等价于满足 $\pi(\mathsf{V}_{\mathsf{H}}^{\mathsf{in}}) = \mathsf{V}_{\mathsf{L}}^{\mathsf{in}}$ 的向下抽象，并且存在一个（唯一的）查询映射 π' 使下图交换。

$$\begin{array}{ccc} \mathsf{Inc}(\mathsf{V}_H) & \xrightarrow{\pi} & \mathsf{Inc}(\mathsf{V}_L) \\ {\scriptstyle |-|_{\mathbb{H}}}\downarrow & & \downarrow{\scriptstyle |-|_{\mathbb{L}}} \\ \mathrm{Do}(\mathsf{V}_H) & \xrightarrow{\pi'} & \mathrm{Do}(\mathsf{V}_L) \end{array} \tag{72}$$

证明。 我们验证 (26) 等价于 (72) 中存在 π'。由于 $|-|_{\mathbb{H}}, |-|_{\mathbb{L}}$ 保留所有变量，因此对所有 V 都有$\pi'(\mathrm{V}) = \pi(\mathrm{V}) \in \mathsf{V}_H$。正如定理25的证明中所述，任何保留输入的查询映射 π' 都必须将每个 $\mathrm{Do}(S)$ 映射到 $\mathrm{Do}(\pi'(S))$，因为后者是唯一具有正确输入和输出的查询（即将 (58) 中的 π 替换为 π'即可定义）。现在，给定这个唯一映射 π'，可以看出该图交换当且仅当 (26) 成立。 □

引理 80. 对于任意定义在E上的 *FCM* $\mathbb{M}$，U 定义一个抽象查询集合$\mathrm{Do}(\mathrm{E},\mathrm{U})_\lambda$，它由$\mathrm{Do}(\mathrm{E} \cup \{\mathrm{U}\})$的不交并构成，其中将单一类型U 视为输入，此外还包含一个由U的状态给出的额外查询λ。将每个查询映射到其图表因子的函子 $|-|: \mathcal{L}^3(\mathrm{E}) \to \mathbf{S}$ 可通过函子分解$|-| : \mathrm{Do}(\mathrm{E},\mathrm{U})_\lambda \to \mathbf{S}$，我们将该函子记作 $|-|$。反事实抽象恰好是一个向下抽象 $(\mathbb{L}, \mathcal{L}^3(E_L)) \underset{\pi}{\overset{\tau}{\rightleftarrows}} (\mathbb{H}, \mathcal{L}^3(E_H))$，使得存在一个（唯一的）查询映射 π'，令下图交换：

$$\begin{array}{ccc} \mathcal{L}^3(E_H) & \xrightarrow{\pi} & \mathcal{L}^3(E_L) \\ {\scriptstyle |-|_{\mathbb{H}}}\downarrow & & \downarrow{\scriptstyle |-|_{\mathbb{L}}} \\ \mathrm{Do}(E_H, \mathsf{U}_H)_\lambda & \xrightarrow{\pi'} & \mathrm{Do}(E_L, \mathsf{U}_L)_\lambda \end{array} \tag{73}$$

和 $\pi'(\mathsf{U}_H) = \mathsf{U}_L$。

证明。 我们来验证，这与向下抽象采用形式（32）等价。对于变量 $\mathrm{X} \in E_H$，根据构造，我们有 $\pi'(\mathrm{X}) = \pi(\mathrm{X})$。与定理25的证明一样，为了保持类型，任何保持输入的查询映射 π' 都必须将每个 $\mathrm{Do}(S)$ 映射到 $\mathrm{Do}(\pi(S))$，并且根据构造，必须将 λ 映射到 λ。因此，π' 由 π 唯一确定。于是可以看出，（73）可交换当且仅当（32）成立。事实上，将每个查询 $Y_1|_{S_1}$、…、$Y_m|_{S_m}$ 对应的图表写出后，应用 π' 意味着对 $\mathbb{F}$的每个开口方框应用 π，这恰好得到查询 $\pi(Y_1)|_{\pi(S_1)}$、…、$\pi(Y_m)|_{\pi(S_m)}$对应的图表。 □

[56]注意，Inc(V) 的每个生成元确实都可以映射到自由范畴 Do(V)中的一个图表，因此这样的函子存在。

F.3 范畴 **模型**(C)

从范畴论视角来看，有一种简洁的方式来表示并概括本文的主要定义；不过，为求简洁，前文省略了这种方式。现在我们简要介绍这一方法。下文固定一个 d-范畴或 cd-范畴 $\mathbf{C}$，其中所有函子和自然变换均属于相应的适当类型。

定义 81。 我们将 **模型**($\mathbf{C}$) 记作这样一个范畴：其对象 $\mathcal{G}^{\mathbb{M}}$ 是由一个签名 $\mathcal{G}$ 以及 $\mathbf{C}$ 中的一个 $\mathcal{G}$模型 $\mathbb{M}$ 组成的二元组。对于一个态射 $(F,\ \tau)\colon\ \mathcal{G}^{\mathbb{M}} \to \mathcal{G}'^{\mathbb{M}'}$，我们指的是一个函子 $F\colon\ \mathbf{G} \to \mathbf{G}'$ 以及如下的一个满自然变换 τ。

$$\begin{array}{ccc} \mathbf{G} & \xrightarrow{\ F\ } & \mathbf{G}' \\ & {\scriptstyle [\![-]\!]} \swarrow \quad \overset{\tau}{\Longleftarrow} \quad \swarrow {\scriptstyle [\![-]\!]'} & \\ & \mathbf{C} & \end{array}$$

复合定义为 $(G,\ \tau') \circ (F,\ \tau) = (G \circ F,\ \tau \circ \tau'_F)$，其中 $\mathrm{id}_{\mathcal{S}} = \mathrm{id}, \mathrm{id}$。

简而言之，我们通常用 F代替 (F, τ^F) 来表示这样的态射。我们称一个态射 F为:

- 严格的，当 $\tau = \mathrm{id}$，从而 $[\![-]\!]' \circ F = [\![-]\!]$。由此很容易看出，严格态射在复合下封闭。

- 当映射 F 将生成元映射到生成元时，称其为.....；当每个$\mathbf{G}'$ 的生成元都可表示为 $F(\mathsf{g})$，其中 g 是 $\mathbf{G}$的某个生成元时，称其为满射 。

在$\mathbf{C}$ 中的任何结构 $\mathcal{S}$ 与查询 $\mathcal{Q}$ 的组合模型 $\mathbb{M}$ 都对应于$\mathbf{Model}(\mathbf{C}${v13>中的两个不同对象 $\mathcal{S}^{\mathbb{M}}$ 和$\mathcal{Q}^{\mathbb{M}}$ 。于是:

- 当存在严格态射 $|-|\colon\ \mathcal{Q}^{\mathbb{M}} \to \mathcal{S}^{\mathbb{M}}$时，称查询 $\mathcal{Q}$ 是抽象的。
- 向下抽象 $\mathbb{L} \to \mathbb{H}$ 恰好是一个映射 $\pi\colon\ \mathcal{Q}_H^{\mathbb{H}} \to \mathcal{Q}_L^{\mathbb{L}}$。
- 当存在一个态射 π^S 使得下图可交换时，向下抽象是组件级的:

$$\begin{array}{ccc} \mathcal{Q}_H^{\mathbb{H}} & \xrightarrow{\ \pi\ } & \mathcal{Q}_L^{\mathbb{L}} \\ {\scriptstyle |-|}\downarrow & & \downarrow{\scriptstyle |-|} \\ \mathcal{S}_H^{\mathbb{H}} & \xrightarrow{\ \pi^S\ } & \mathcal{S}_L^{\mathbb{L}} \end{array}$$

当 π^S （因此 π也）是严格的时，组件级向下抽象的情形是严格的。

例 82. 存在一个严格映射 $\mathsf{io}(\mathsf{V}) \to \mathrm{Do}(\mathsf{V})$，它在变量上是单射。构造性因果抽象恰好是一个态射 π，使得下图交换。

$$\begin{array}{ccc} \mathsf{io}(\mathsf{V}_H)^{\mathbb{H}} & \overset{\pi}{\dashrightarrow} & \mathsf{io}(\mathsf{V}_L)^{\mathbb{L}} \\ \downarrow & & \downarrow \\ \mathrm{Do}(\mathsf{V}_H)^{\mathbb{H}} & \xrightarrow{\ \pi\ } & \mathrm{Do}(\mathsf{V}_L)^{\mathbb{L}} \end{array}$$

我们可以如下处理向上抽象。将一对映射称为变换:

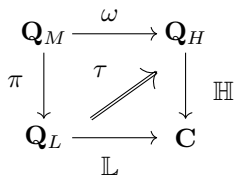
$$\mathcal{Q}_L^{\mathbb{L}} \xleftarrow{\ \pi\ } \mathcal{Q}_M^{\mathbb{M}} \xrightarrow{\ \omega\ } \mathcal{Q}_H^{\mathbb{H}} \tag{74}$$

使得 ω 是严格的、满射的，并且在变量上是恒等映射。那么:

- An 向上抽象是 π 在类型和生成元上为单射的特殊情形 .
- 向下抽象是 ω 在生成元上为双射的特殊情形，因此我们可以将$\mathcal{Q}_M^{\mathbb{M}} = \mathcal{Q}_H^{\mathbb{H}}$视为相同。于是它就是一个映射 $\pi\colon\ \mathcal{Q}_H^{\mathbb{H}} \to \mathcal{Q}_L^{\mathbb{L}}$。

尤其是在处理比本文所关注的更复杂的给定查询与模型之间的关系时，采用范畴$\mathbf{Model}(\mathbf{C})$并使用上述术语可能会很有帮助。我们将这一范畴及其性质的全面探究留待未来工作。

一种更一般的函子观点下的抽象。 从范畴论的角度来看，模型以及模型之间抽象的最一般概念，或许是由若干范畴（不一定带有给定的签名）、函子和一个自然变换构成的集合：



使得这一对 (π, ω) 构成一个 关系 ，即满足 $\langle \pi, \omega \rangle \colon \mathbf{Q}_M \to \mathbf{Q}_H \times \mathbf{Q}_L$ 是忠实的。这意味着，我们可以将抽象理解为一种以结构化方式关联高层级和低层级“查询”的方法。也就是说，对于关系中包含的那些态射（“查询”）(Q_H, Q_L)，我们可以得到典范一致性方程 (34)。允许任意形式的此类关系，可以提供比我们在此考察的方式更为一般的结构化方法，来关联高层级和低层查询（不一定由部分函数给出）；不过，这值得在未来工作中进一步探索。