第 3 章　信念修正与社会影响*

当个体从对自身信念或知识的推理转向对他人信念或知识的推理时，我们需要扩展概念、对更多的情境做出描述和区分。譬如，第 2 章中已经引入的公共知识（每个人都知道且知道她们知道……）和分布式知识（综合所有知识在一起，我们将会知道……）之间的区别。考虑不同的主体聚集在一起这样的一种现象，研究的问题自然也从个体的信念修正延伸到群体信念或群体意见的形成。此外，当我们考虑到社会关系时，许多关于知识传播、信念改变的问题会变得更为有趣。本章主要研究的一个问题是社会影响与认知之间的关系，即，个体在受到与其有社会关系的其他主体的影响下如何改变自己的信念，以及变化的规律是什么。

首先，我们需要声明，关于社会关系的建模不是逻辑学家们的首创。在社会心理学、人工智能和经济学中，社会关系都扮演着重要的角色。例如，文献 [8] 提出了具有开创性的工作，它从社会权力的角度研究群体的人际关系模式，随之产生了许多处理动态更新的数学和计算模型。又如，受市场营销应用的推动，文献 [7] 和 [13] 研究了如何找到社会网络中具有影响力的成员、如何实现社会影响力最大化等问题。社会网络在经济学中也得到了广泛的研究，例如颇具代表性的工作（文献 [12]）。此外，还出现了一个被称为“社会模拟”的新兴领域，该研究领域使用计算模型解释社会中的一些动态现象，并对其未来发展进行预测（参见文献 [9]）。

然而，学者们很少关注如何引导人们思考关于社会关系推理的规范。这也许是西方知识论的个体中心主义导致的结果：其他人天生不可靠，所以主体不可能以回应朋辈观点为纯粹的逻辑依据来改变自己的信念。尽管如此，日常生活中主体确

---

* 本章部分内容首次发表于：Fenrong Liu, Jeremy Seligman, and Patrick Girard, Logical Dynamics of Belief Change in the Community, *Synthese*, 191(11): 2403-2431, 2014. 此次收录本书时进行了调整和修订。

实通过这种方式改变自己的信念。信息沿着社会关系的通道进行传播，这是传统知识论未曾涉略的问题，这对知识论的进一步发展具有重要的意义。需要强调的是，我们的出发点是逻辑学，我们将会研究如何对这种信念修正进行推理，而推理的规律是什么。这一章的工作基于之前的研究（文献 [18]，特别是文献 [15]），探究主体信念改变与社会影响之间的关系。

以下简要说明本章的方法论。首先，我们是沿着动态认知逻辑的传统继续发展（文献 [21]）。动态认知逻辑领域的学者们同样致力于研究信念修正的模型，具有代表性的工作有：文献 [4]，[20] 和 [23]。但与他们的研究不同，本章的视角是将社会关系的角色考虑进来。因此，我们很少涉及信念修正的具体细节，因为这里提出的工作与许多此类研究和解释是一致的。其次，本章引入很多例子，以轻松的风格写成，希望既能面向广大读者，也能让我们有空间去涵盖更广泛的主题。我们希望这种方式能够展示在研究信念变化逻辑时，认真对待社会关系是十分有益的。对于那些期望技术细节的读者，建议参考本书的其他章节或相关的论文，如文献 [10]，[18] 和 [19]。最后，因为关注的重点不同，我们不会使用概率或其他定量计算方法研究社会信念的变化。这里的目标是研究信念变化的逻辑，即如何推理主体在受到他人影响时改变自己信念的方式。定量方法与该研究课题存在某种关系，我们将在后面的章节继续探讨。在逻辑的定性方法和定量方法之间建立进一步的桥梁显然是一个有价值的课题，本书很多地方将会涉及此问题。

本章的结构和内容概述如下：在 3.1 节我们首先介绍信念修正的标准模型，指出该模型在文献 [1] 的传统中可被扩展至个体信念的社会影响模型。特别是，在 3.1.2 节中我们介绍一种基于阈值假设的特定模型，要求主体在受到他人影响时改变自己的信念。这种模型可以用来展示各种信念变化的动态现象，并可以使用自动机的方法进行逻辑建模和分析（见文献 [15]）。在 3.2 节，我们进一步研究使群体信念在社会影响下保持稳定的原因。同时，我们会探讨一些动态变化的具体情境，如某个体单边改变自己的信念（3.2.1 节），新的社会关系的形成或消解（3.2.3 节）。我们还会考虑这些变化对整个群体信念聚合的影响（3.2.2 节）。3.3 节介绍其他可能的模型。例如，考虑引入主体对其他人的排序，可以根据他人可靠程度的高低进行排序。最后，3.4 节从本章所依赖并继承的单命题修正/收缩模型本身出发，企图通过基于合理性关系的模型（来自文献 [6] 和 [24]）提供一种更一般的研究方案。

3.1　社会影响下的信念变化

所谓“受到朋友的影响”就是这样一个动态过程：一个主体改变自己的信念，使其更符合周遭朋友的信念。以下考虑关于单命题 p 的影响。如果我不相信 p，而我的朋友中有相当一部分人相信它，我可以用以下几种方式来回应。当然，我可以无视她们的意见，镇定自若。但如果我确实受到影响并改变了自己的信念，存在至少两种达成方法：

第一，我可能会修正信念使自己也相信 p；

第二，更谨慎一些，我可能收缩信念，即，仅仅消除我对其否定 $\neg p$ 的信念。

我们将修正行为记作 Rp，收缩行为记作 Cp。[①] 我们对修正和收缩的唯一假设条件是它们是“成功的”。这意味着当我执行 Rp 后，我将相信 p，并且，当我执行 Cp 后，我将不会相信 $\neg p$。[②] 在逻辑上，这意味着接受以下公理：

$$[R\varphi]B\varphi$$
$$[C\varphi]\neg B\neg\varphi$$

主体是否会改变自身信念以回应朋友们的意见？如果改变，是修正还是仅仅收缩？这里的决策取决于至少三个因素：

（1）主体关于 p 的态度；

（2）朋友们关于 p 的信念的内聚力；

（3）在多大程度上，主体将某特定朋友作为对 p 判断的权威。

为简化问题，我们先假设所有人都认为她们的朋友在所考虑的每一个问题上都具有同等权威。这个假设将在 3.3.1 节和 3.3.2 节中被取消。面对的异议内聚力越大，主体就愈有压力去改变。但同时，如果主体一开始就对 p 持开放的态度，可能会比持强烈反对意见时更容易受到朋辈的影响。这些因素的具体平衡因人而异，

① 任何信念修正理论的支持者都可以根据自己喜欢的理论来解读 Rp 和 Cp。上一章介绍的 AGM 解释足以应对我们的目的，但此处讨论不会过分依赖于细节。

② 虽然成功条件被许多关于信念变化的理论作为基本假设，包括 AGM 理论，但它确实施加了一些限制。特别是许多高阶命题，如摩尔类的命题形式“p 但我不相信 p”，都是有问题的。

甚至是因信念而异。它也可能部分地由主体对朋友或其他事情的可靠性的更高层次信念所决定。因此，我们不拘泥于一种关于影响的特殊理论，而只是区分两种影响: 导致修正的影响和导致收缩的影响。在主体受影响去积极地修正自身信念以支持 p 的情况下，我们说主体*受强影响而相信* p，并将其记作 $\mathsf{S}p$。可能还有其他原因使得主体的信念改变，但仅就社会影响而言，且其他条件不变的情况下，条件 $\mathsf{S}p$ 是主体改变信念的必要和充分条件。同样地，当主体受到影响而仅仅收缩其关于 $\neg p$ 的信念（如果有的话），而不一定相信 p 时，我们说主体*受弱影响而相信* p，并记作 $\mathsf{W}p$。我们还将涉及相应的受强、弱影响去相信 $\neg p$ 的消极条件，分别记作 $\mathsf{S}\neg p$ 和 $\mathsf{W}\neg p$；此外，我们将假设不可能同时受强或弱影响去既相信 p 又相信 $\neg p$。

引入这些术语和符号后，我们就可以定义关于 p 的社会影响的一般动态行为。主体受关于 p 的影响，记作 Ip，指的是主体在强影响下修正自身信念以相信 p；否则，在弱影响下收缩其为 $\neg p$ 的信念（如果有的话），$\neg p$ 的情形类似。如果主体连弱影响都未经受，则其信念保持不变。更确切地说，Ip 可以被定义为如下程序：[①]

$$\text{若 } \mathsf{S}p \text{ 则 } Rp\text{，否则，若 } \mathsf{W}p \text{ 则 } C\neg p\text{；}$$
$$\text{若 } \mathsf{S}\neg p \text{ 则 } R\neg p\text{，否则，若 } \mathsf{W}\neg p \text{ 则 } Cp\text{。}$$

从逻辑学的角度来看，这意味着使用命题动态逻辑（Propositional Dynamic Logic, 简记为 PDL）[②]的方法处理，可以很容易地将 R、C、S 和 W 的逻辑系统扩展到 I 的系统。尽管本章不会探讨技术层面的逻辑问题，但这一特性将允许我们在模态语言背景下根据无迭代PDL表达式的已知归约的结果，对社会信念变化的动态特性表述进行归约。关于PDL的公理化，参见如文献 [5] 中的定义 4.78 或者文献 [11]。关于动态认知逻辑（如文献 [3] 和 [22]）中动态算子的更一般化的研究，可以参阅论文 [10]。

在本书后面的章节，我们将提供各种不同的强、弱影响的具体描述，进一步给出形式语言，其中包含群体的信念和社会关系的算子。不关心技术细节的读者可以忽略这些形式化的讨论，不会影响本章的阅读。

---

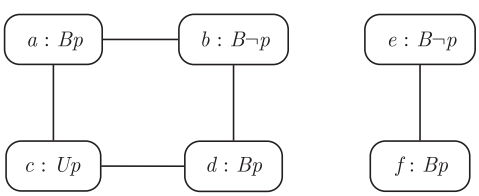
① 在我们关于一个主体不能同时既受影响去相信 p 又受影响去相信 $\neg p$ 的假设下，肯定或否定子句的顺序并不重要。

② PDL 是模态逻辑的扩展，将每个程序看作一个模态算子，可能世界语义模型扩展为状态转换系统，程序的执行导致状态之间的转换。具体细节，可参考文献 [11]。

3.1.1　影响的动态

关于命题 p，主体有三种可能的信念状态: 相信（Bp）、不相信（$B\neg p$）[①] 和没有信念 $(\neg Bp \wedge \neg B\neg p)$，缩写为 Up。为了讨论这些状态在朋友之间的分布，我们使用在文献 [18] 中引入的“群体逻辑”框架。在该框架中，友谊或朋友关系被看作一种对称且非自反的关系。这就是说，我是我任何朋友的朋友（对称性），而我不是我自己的朋友（非自反性）。友谊不是传递的，所以，很有可能我的朋友的朋友不是我的朋友。友谊所关联的主体的集合将被称为社会网络。通过友谊所连接的主体的子集被称为社群，此处“连接”意为该子集中任意两个主体都能通过朋友关系联系起来。

以下给出一个实例，展示社会网络中信念状态的分布:

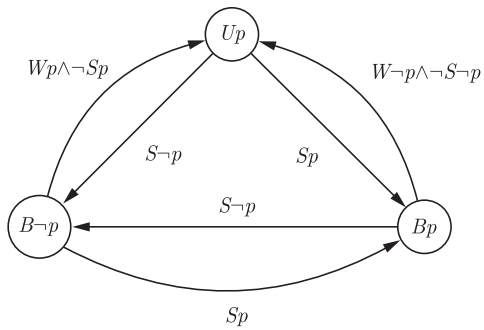


上图代表了一个由六个主体所组成的社会网络，聚集成两个社群。主体 a 相信 p，并且 a 有 b 和 c 两个朋友; 主体 b 不相信 p，并且 b 有 a 和 d 两个朋友, 等等。除了描述主体的信念状态（Bp，$B\neg p$，或 Up），我们还将使用符号 F 表示“我所有的朋友”, 来描述她们在社交网络中的位置。因此，FBp 意味着我所有的朋友都相信 p。在上面的例子中, 这是对主体 b, c, 和 e 的真实描述, 但对于主体 a, d 和 f 该命题不成立。其对偶形式 $\langle F\rangle$ 表示“我的一些朋友”。例如，$\langle F\rangle Up$ 表示我至少有一个朋友对 p 的信念态度尚未确定。这在图中仅适用于主体 a 和 d。

给定修正和收缩的成功条件的公理，每个主体的信念状态将在这些操作下以确定的方式进行改变。$\neg p$ 的修正将使主体从状态 Bp 转换到状态 $B\neg p$，其收缩则使主体从 Bp 转换到 Up, 等等。此外，如果我们假设强影响和弱影响的触发条件仅取决于社会网络中各主体的信念状态分布，则在操作 Ip 时，这种分布将以完全

① 注意 $\neg Bp$ 与 $B\neg p$ 之间的区别。

确定的、局部的方式发生变化。社会影响的动态变化（对每个主体的、局部的）可以由以下有限状态自动机所刻画: ①



自动机的状态是主体可能的信念状态。转换的条件用相互排斥的影响条件来标记。例如, 如果某主体相信 p, 由标记着 Bp 的状态表示，并且受她的朋友弱（不是强）影响去相信 $\neg p$，用标记 $(\mathsf{W}\neg p \wedge \neg \mathsf{S}\neg p)$ 的状态表示，之后她会收缩她关于 p 的信念, 并处在状态 Up 上，即对 p 的态度保持中立。

这一自动机模型以一种十分简约的方式，对主体受到周遭朋友的影响而发生认知状态的转换进行建模，为我们理解群体的信念修正提供了一种普遍的框架，也为我们进一步讨论问题提供了可能。

3.1.2 阈值模型

为说明社会影响的具体内涵，我们给强影响和弱影响以更为明确的解释。第一种方案是，强影响需要在一定阈值比例的朋友间形成内聚。保守而言，我们将阈值假定为 100%，这意味着我受强影响去相信 p，当且仅当，我所有的朋友都相信 p（且我至少有一个朋友相信 p）。括号里的句子使得“所有的朋友”有实际的意义，

① 通过有限状态自动机以分析社交逻辑动态最初由文献 [15] 提出，其表明一些有趣的动态属性（比如最终收敛到一个稳定的分布）可以被类似于我们这里所考虑的算子所表示，不过该论文中讨论的是偏好而非信念。这里我们提供了一种对信念变化的更一般的刻画，它不依赖于任何关于修正、收缩、强或弱影响的具体描述。就像在文献 [15] 中那样，重要的是要意识到自动机不是动态系统的定义，而是一种用来分析那些隐含在逻辑算子定义中的对象的工具。

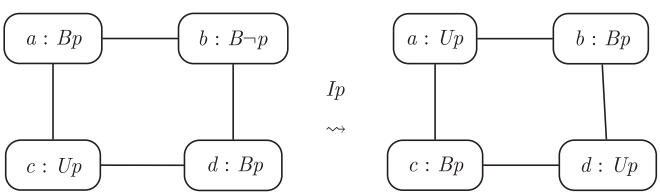
否则，就逻辑学的量词而言，没有朋友相信 p 时，“我所有的朋友都相信 p”句子也是成立的。[①] 受强影响去相信 $\neg p$ 的过程类似。就弱影响而言，假设一个人至少有一个朋友相信，并且不相信的朋友数量不超过某个阈值。同样，我们可以假设阈值是 0。这意味着，如果我没有受到强影响去相信 p，那么我受弱影响，当且仅当，我的朋友中没有人相信 $\neg p$ （且我至少有一个朋友相信 p）。这种对强影响和弱影响的解释可以用以下公理来表述:

$$S\varphi \leftrightarrow (FB\varphi \wedge \langle F\rangle B\varphi)$$

$$W\varphi \leftrightarrow (F\neg B\neg\varphi \wedge \langle F\rangle B\varphi)$$

下面，我们考虑一个单步影响的实例：

例 1（单步的影响）



在这个例子中，在刚开始的时候，主体 a 和 d 都相信 p，后受弱影响而放弃这个信念，因为她们所有的朋友都不相信 p （b 不相信，c 不确定），且她们的一个朋友 b 相信 $\neg p$。因此对 a 和 d 而言，$W\neg p$ 是真的，并且在社会影响 Ip 的作用下，她们都收缩了其对 p 的信念而变得犹豫不决。相反，主体 b 和 c 会受到其朋友 a 和 d 的强影响而相信 p。因此，Sp 对于 b 和 c 而言是真的，并且在 Ip 的作用下，她们都进行了修正，从而相信 p 。注意，这里我们假设所有主体在同一时间修正她们的信念。另一种自然的情境是每个主体按一定顺序修正自己的信念，对应于各主体在不同时间基于他人信念进行修正。若对这样的情形进行研究，需要使用时态逻辑，表达在不同的时间点主体的认知状态及其变化，这似乎是一个有趣的研究方向。

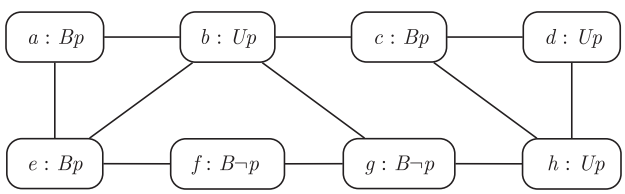
① 对全称量化句的解释，在一阶逻辑教材中容易找到。

3.2 稳定与变动状态

例 1 展示了 Ip 算子单次运用后所发生的情况。从动态变化的过程可以看到，进一步的改变仍会发生。在右边的认知分布中，a 和 d 都受强影响去相信 p，因此进一步运用 Ip 会使她们改变自己的信念，从而回到先前的状态。如果继续下去，会发生什么呢？在这种情况下，主体 b 和 c 没有进一步的变化，所以再次运用 Ip 后，所有人都一致地相信 p。社会影响将不会引发进一步的变化。

如果算子 Ip 对社群中任何主体的信念状态都没有进一步的影响，我们就称该社群是*稳定的*。社群内的全体成员具有一致的信念是维持社群稳定的充分而非必要条件，如例 2 所示：

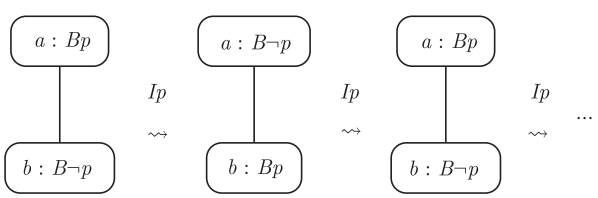
例 2（稳定社群）



在这个仅由单一社群组成的社会网络中，没有主体会受到强影响或弱影响。例如，主体 b 对 p 犹豫不决，她有三个朋友相信 p，但有一个不相信。根据我们对阈值影响给出的假设，这不足以让她改变信念。还应注意的是，主体 c 的朋友们对 p 不置可否的信念是一致的，但这对 c 关于 p 的信念没有影响。

例 1 中的社群刚开始不稳定，但运用一次 Ip 后就变得稳定了。我们将这样的社群叫作“*将要稳定*”的社群。并非所有社群都会变得稳定。那些永远不稳定的社群将被称为“*变动着*”的社群。试举一例：

例 3（变动）



在例 3 中，各主体在每一阶段都受到对方的强影响，并因此修正她对 p 的信念，在相信和不相信 p 间交替变化。

以上的讨论表明，就社群的状态而言，有三种可能的社群类型: 稳定的社群、变动的社群和目前不稳定却将要稳定的社群。此外，由于一个社会网络中不同社群彼此间没有影响，因而同时包含几种不同类型社群的社会网络完全是有可能的。

用逻辑的形式语言描述稳定性是比较容易的，因为它可以在局部完成。一个主体如果不受强影响或弱影响而相信 p，也不受强影响或弱影响而相信 $\neg p$，是不会改变她的信念的。同时，对于已经相信 p 的主体，在受到强影响或弱影响而相信 p 时，也将保持信念不变。回到自动机模型, 容易看到，以下条件对于一个主体的信念不变是必要且充分的（假设强影响蕴涵弱影响）:

$$\neg(B\neg p \wedge \mathsf{W}p) \ \wedge \ \neg(Up \wedge \mathsf{S}p) \ \wedge \ \neg(Up \wedge \mathsf{S}\neg p) \ \wedge \ \neg(Bp \wedge \mathsf{W}\neg p)$$

以上的四个子句分别描述四种情形：主体将停留在自动机的一个状态上, 且不会改变她的信念状态。在阈值影响的假设下，使用朋友关系和信念算子，这可以进一步表示为:

$$\neg(B\neg p \wedge F\neg B\neg p \wedge \langle F\rangle Bp) \wedge$$

$$\neg(\neg Bp \wedge \neg B\neg p \wedge FBp \wedge \langle F\rangle Bp) \wedge$$

$$\neg(\neg Bp \wedge \neg B\neg p \wedge FB\neg p \wedge \langle F\rangle B\neg p) \wedge$$

$$\neg(Bp \wedge F\neg Bp \wedge \langle F\rangle B\neg p)$$

当社群中每个主体都满足这一条件时，社群就是稳定的。刻画变动（或将要稳定）的状态有点困难。文献 [15] 包含了一个定理，展示了如何为偏好动态变化给出类似解释。这里，我们尝试给出关于信念一个类似的猜想，即一个社群（有至少两个主体）是变动的，当且仅当，社群中的每一个主体都满足如下条件:

$$(FBp \wedge FFB\neg p) \ \vee \ (FB\neg p \wedge FFBp)$$

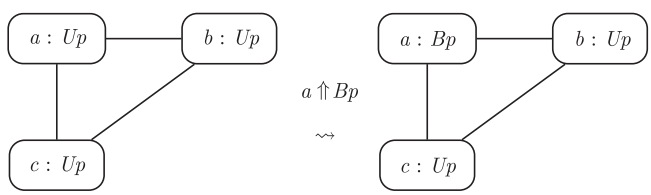
特别地，如果社群中有任何主体在状态 Up 上，且该社群尚未稳定的话，那么它将要达到稳定状态。要想用形式语言来描述它，我们需要一些全局的视角。最终，在有限次数的更新之后（可以引入 $[Ip]^n$），保持稳定的条件对每个主体都有效。

3.2.1 节中，我们将研究如何能够打破社群的稳定性，或者如何能够通过一些方法促进一个社群走向稳定。当然，我们既考虑信念在一个社群内部传播的情况，也考虑信念在社群之间的传播。

3.2.1 单边信念改变

抛开朋友意见的影响外，主体可能会因很多其他原因改变她们的想法。这就自然提出了如下问题: 这种变化能否以及如何传播到社群的其他成员？一个非常团结的社群可能会抵制所有这些可能的变化，能够确保任何单边信念改变的主体很快恢复原来的信念。而一个不那么团结的社群，可能会受这种变化的高度影响而开始发生变化，甚至是跟随改变信念的主体进入一个新的稳定状态。我们将在这里考察一些例子，从一个全体成员一致中立的社群中单个主体转而相信 p 开始：

例 4（接受新信息）



在最初的稳定社群中有三个主体，她们都是朋友，而且都没有确定的信念。现在，主体 a 改变她的信念为 Bp，在图中用“$a \Uparrow Bp$”表示。然而，这个变化非常有限，因为生成的新状态也是稳定的。在信念改变后，主体 a 没有放弃自身信念的压力，因为她所有朋友的信念状态仅仅是中立。那些朋友，b 和 c，处在相信 p 的弱影响下，在阈值影响解释下，这还不足以使她们改变自己的中立状态。因此，主体 a 的单边信念改变完全被孤立。

有时，社群内的主体的一致信念可以强大到足以抵制单边信念改变，下面的例 5 展示了这一点。