

行动性，因果与意图¹

费定舟

(武汉大学哲学系，武汉，430072)

摘要：如何描述行动性 (agency) 是行动哲学的中心课题，寻找行动性的哲学分析与逻辑表达是许多研究者的重要任务。本文概括了有关行动性的主要理论，特别讨论它与因果性的关系，以及相应的逻辑表达形式。在此基础上，我们指出了其中存在的困难与不足之外。为此，一个改进的尝试方案被提出：它以 Searle 的行动中的意图为哲学基础，以动态逻辑为主要逻辑表达手段。此外，本文还比较了我们的方案与其它行动性理论的异同，同时指出了值得进一步探讨的方向。

关键词：Agency 树框架 事先意图 Stit 理论 行动中的意图

中图与类号：B81 **文献标识码：**A

行动 (action) 与行动性 (agency) 是行动哲学 (philosophy of action) 的中心论题 [R. Mele, 1997]。什么是行动？行动如何解释？在大部分情形下，哲学家们用富于心理内容的词汇来描述，刻画和定义这些问题，这种讨论经常与诸如原因 (Reason)，因果性 (Causality) 等概念联系在一起 [J. Hornsby, 1993]。

但是从逻辑层面上讨论，则是另外一种处理方式了。也有许多哲学家，尤其是逻辑学家感兴趣的是把行动性作为模态逻辑的应用课题，例如，行动性的概念，Agent 在行动的作用和地位等探讨被看作行动逻辑 (logic of action) 的主要内容 [R. Hilpinen, 1997]。

我在本文中将主要遵循后一种研究方法的路线，着重从逻辑形式上讨论进行逻辑中这两个核心课题，尽管如此，我们也需要前一种处理方式提供的概念基础。

在关于行动性的逻辑研究中，有两种传统可以辨认 [N. Belnap, 1991]。其中一种讨论 Agent 引起 (brings it about) 某种事态的发生的句子，类似于一个模态算子作用于这种事态的公式。Anselm, S. Kanger, M. Brown, N. Belnap 等是其中的代表。另一种传统把行动句只看作某个 Agent 的做事 (doing things) 而已，就象 Von Wright 提倡的那样 [Von Wright, 1963]。在不同作者的研究中，这两种传统实际上是在不同的程度上混合起来了。

我在本文中将挑选几个到目前为止有代表性的关于行动性的理论和逻辑形式，阐述它们

的形式特征与直观涵义，讨论它们面临的一些困难。并且，本文将提出尝试解决这些问题的不同办法，并讨论体现在方案背后的哲学基础与相应的逻辑工具。我们的解决办法也是上述两种传统的结合。

本文的结构安排如下：首先是各种代表性理论关于行动性的探讨，接着是它们面临的不利之处，接着，一个基于意图的行动性的逻辑形式与语义模型给出并以回答这些问题，并就有关有待深入讨论的课题进行了讨论。

行动性：肯定与否定方面

行动性理论的核心问题是这样表述的：

(a) i 是行动 $a_1, a_2 \dots$ 的 Agent 当且仅当如此这般 (such and such) 或

(b) i 引起 (bring it about, 或 sees to it) 事态 P 当且仅当如此这般

(a) 代表了将行动性看作程序的立场，它属于 Von Wright 所倡导的传统；(b) 代表了将行动性看成了类似模态算子之类的作用于一阶公式的算子，属于 Anselm 所倡导的传统。比较一下这两种表示是有趣的：Belnap 把 Agent 看成所有动作的集合²，因而 (b) 成为同语反复而无甚意义，而 (a) 则看不出动作的重要性，它强调引起的后果。

在这部分我们概括出行动性的两个不同的方面：其中一个指出某个事态 (the state of affair) 成立的条件，另外一个则指出它不成立的条件。在不同的逻辑学家那里，具体称呼是不一样的。但都表明，要完全刻画行动性需要两个不同的角度来进行。S.Kanger 曾定义了一个行动算子 $Do_i P$ (i see to it that p) [Kanger, S, 1972]:

(1) $Do_i P \quad D\acute{o}_i P \quad D\grave{o}_i P$

这里 $D\acute{o}_i P, D\grave{o}_i P$ 解释如下：

(2) $D\acute{o}_i P = P$ 是充分的，对于 i 所做之事而言；

(3) $D\grave{o}_i P = P$ 是必要的，对于 i 所做之事而言。

假定 $u, w \in W$ 是非空集合 W (称为可能世界) 的元素， $R_i(u, w), S_i(u, w)$ 分别是 u, w 之间的可通达关系， $R_i(u, w)$ 描述充分条件下的行动性， $S_i(u, w)$ 描述必要条件下的行动性，即

(4) $u \models D\delta_i P$ 当且仅当对于所有满足 $R_i(u, w)$ 的 w 而言, $w \models P$ 。

(5) $u \models D\delta_i P$ 当且仅当对于所有满足 $S_i(u, w)$ 的 w 而言, $w \models \neg P$ 。

$R_i(u, w)$ 意指 u, w 中, *agent* 以相同方式行动 (B. Challas, 1969), 这样得到的语义模型与模态系统 T 相当, 即 $R_i(u, w)$ 是自反的。 $S_i(u, w)$ 则指 u 在 w 中 *agent* 所做的对立行动 (opposite action), 不过这里 $R_i(u, w)$, $S_i(u, w)$ 的直观含义不是很清晰的。后来, I. Pörn 在某种程度上弱化了上述条件 [I. Pörn, 1977], 即用 $\neg D\delta_i \neg P$ 代替 $D\delta_i P$, 于是

(6) $u \models \neg D\delta_i \neg P$ $u \models Co_i P$ 当且仅当存在 w 使 $S_i(u, w)$, 也使得 $w \models \neg P$, 这里, 定义可解读为: 对于 *agent* 的行动而言, 也许并不总是 P 成立, 这意味着 P 并不是不可避免的。这样定义的直观比 Kanger 的要清楚多了, 它告诉我们, 必要条件是指也许语句 (might-statement), 而不是必须一语句 (would-statement)。Belnap 把 $D\delta_i P$ 与 $Co_i P$ 分别称为行动性的“肯定方面”与“否定方面” (positive, negative condition) [N, Belnap, 1991]。

对行动性的两重分析同样也体现在基于 Von Wright 关于行动的理论之上的逻辑分析之中 [R. Hilpinen, 1997]。我们说过, Von Wright 的工作着重于 Agent 与动作之间的归属关系, Agent 所处周围环境或世界的角色对于行动性来说是十分重要的。有三种世界或情景³: 初始状态 (initial state), 它对应于行动开始之前与世界无作用的状态; 由于行动而导致世界或状态发生了变化, 此时所得到的世界或状态的改变, 称为结果状态 (end-state 或 result state); 反状态 (counter-state)⁴, 从初始状态出发, 但并不需要行动或来自于周围环境的干预, 换句话说, 来源于 Agent 的被动性 (Passivity)。而结果状态则称之为来源于 Agent 的主动性 (activity)。事实上, 依 Von Wright 的观点, 反状态抓住了行动性中的“反事实”因素 (Counter-factual element), 这与 D. Lewis 所言反事实条件句有密切的关系 [D. Lewis, 1973]。记 u 为初始状态, $e(u)$, $d(u)$ 为以 u 为初始状态的结果状态和以 *agent* 的被动性为基础所导致的反状态 $d(u)$ 。只要 *agent* a 历经初始状态经行动 a 到达结果状态后, 我们可以说 Agent 是该行动的制造者 (maker)。考虑任一事态 P , 则 Von Wright 的行动性可用两式定义:

(7) $w \models B_i^* P$ 当且仅当 $w \models P$, 且存在 $u \in W$, 使得 $u \models \neg P$, $w = d(u)$, $e(u) \models \neg P$

(8) $w \models B_i^{**} P$ 当且仅当 $w \models P$, 且存在 $u \in W$, 使得 $u \models P$, $w = d(u)$, $e(u) \models \neg P$

满足 (7) 式所界定的行动性称之为产生型行动性 (productive agency), 满足 (8) 式所

界定的行动性可称为保持型行动性 (preserver agency)。从定义看出,若 Agent a 所相关的行动性是产生型,则它不是保持型,反之亦然。我们亦可以把 (7)(8) 两式用另一种形式写出来。记 $f(a,u)$ 记 Agent i 经 $a \in A$ 行动从初始状态 u 到达的结果状态, A 是 agent i 所执行动作的集合,记 $n(u) = f(\text{Null}, u)$, Null 称为“零行动”(null action),它产生于 Agent a 的被动性,则 (7) (8) 式所表达的产生性与保持性两个不同的行动性定义如下:

(7) $w \models B_i^* P$ 当且仅当存在 $u \in W, a \in A$, 使得 $w \in |P| \cap f(a, u)$, 且 $n(u) \subseteq |\neg P|$,
 $u \models \neg P$.

(8) $w \models B_i^{**} P$ 当且仅当 $u \in W, a \in A$, 使得 $w \in |P| \cap f(a, u)$, $n(u) \subseteq |\neg P|$ 。

这里 $|P|$ 是这样定义的: $|P| = \{w \in W : w \models p\}$ 。实际上,行动性的二重区分是基于 agent 的关于行动 a 的因果性标准[D, Lewis, 1986]。

(9)(i) 如果 agent i 已经执行动作 a, 则定会有事态 P 成立 (肯定)

(ii) 如果 Agent i 不执行动作 a, 则不会有事态 P 成立 (否定)

假如 $w \in f(a, u)$, $w \models p$, 这正是 (i) 所表达的; 若 $w \in f(a, u)$, $w \models P$, 但是 agent i 不执行动作 $a \in A$, 则 $f(\neg a, u) \subseteq |\neg P|$, 这正是 (ii) 所表达的, 注意 $\neg a$ 是指 Agent 不执行 a 这个动作, 它可以看作是另外一个不同的动作, 这与一阶逻辑中的否定是不相同的。

由上面分析,我们得知:行动性的分析是以因果性的分析为基础的;而且在可能世界语义学框架内有相应的语义模型。另外一点是十分重要的:我们可以行动性与因果性均可以在决定论的哲学框架加以描述与理解。相反,如果考虑到非决定论的哲学框架,行动性与因果性就以不同的形式出现。值得注意的是,尽管如此,在描述行动性时,虽然采取非决定论立场,但行动性仍然包含二重性,而否定方面仍然来自于表示行动性的反事实成份。

N.Belnap: Stit 理论

上面提到的行动性是基于决定论的。一旦考虑非决定论下的行动性,非因果的选择是十分重要的。它可以体现为 agent 意图:选择某种历史而不选择另外的历史。在这方面的探索是 N.Belnap 及其合作者所进行的[Belnap. N. Perloff. M, Xu M, 2001]。下面以 Belnap 的 Stit

理论 (FF) 来分析行动性与 agent 选择的意图过程[F. Horty, N. Belnap, 1995]。Stit_i 是 “i sees to it that” 的缩写。例如 “Peter drives his car to work”, 可写为 Will [Peter Stit_p] 或 Stit_i;P. Belnap 的记法有前缀 “Will” 表明这是面向将来的选择, 而这预示他采用非决定论来处理行动性与因果性概念。为什么采用非决定论? 因为我们的世界在面对未来时有多种选择⁵, agent 以选择开始行动, 但选择时可以有不同的行动可供参考, 不同的行动导致的结果也不一样。“stit_iP” 为真 (在 $w \in W$ 处) 当且仅当 P 在 W 处的真由 agent i 的事先选择 (prior Choice) 保证, 或者说, P 在 W 处的真是由 i 的行动性来决定的。

定义 stit_iP 的语义模型与 Von Wright 的有所不同。此时一个模型框架是 $\langle W, \leq \rangle$, 其中 W 是非空集合, 它的元素称为时刻, 可能世界或情境, 而 \leq 是一偏序关系, 但是还有下列限定:

(10) 线性要求: 对于 $m, m', m'' \in W$, 有

$$m' \leq m \quad m'' \leq m \Rightarrow m' \leq m'' \quad m'' \leq m'$$

(11) 相关要求: 对于 $m, m' \in W$, 存在 m'' , 使得

$$m'' \leq m \quad m'' \leq m'$$

(10) 的要求保证了 agent 有一个线性的过去; (11) 保证任意时刻有一个起点或过去。满足 (10) 及 (11) 的 Kripke 框架有时称为一个树框架。为了给出语义模型, 需要引进几个重要概念。

历史 (history): W 的子集, 一个极大的线性链 (依 \leq), 记为 h。对于 $m \in W$, 定义 m 处的历史:

$$H_m = \{h : m \in h\}$$

$i \in \text{Agent}$, i 是有选择能力的个体, 称为 agent,

瞬间 (Instant) 把树分成等价类, 记为 Instant, 它是非空集合, 若 $\text{instant} \in \text{Instant}$, 则 $\text{instant} \cap h$ 是单点集。直觉上, instant 表示不同历史的同一时间。

选择函数 Choice_i^m , 它是一个到内映射, 即

Choice: Agent $\times \{m\} \rightarrow P(H_m)$, $P(H_m)$ 为 H_m 的幂集, 记为 Choice_i^m 。

直观上, Choice_i^m 是对 H_m 一个分割(等价类的集合)。特别地, 若 $h \in H_m$, 则 $\text{Choice}_i^m(h)$ 表示 Choice_i^m 中包含 h 作为元素。

选择等价: (Choice_i^w - equivalent), 如果 $m_1, m_2 \in W$ 属于同一 instant, 且

如果 $m_1, m_2 \in W$ 属于同一 instant, 且 $w < m_1, w < m_2, m_1$ 与 m_2 落在这些历史上: 这些历史是 Choice_i^w 的元素, 这些历史的集合包括与 m_1, m_2 相应的 H_{m_1}, H_{m_2} 。于是这两个时刻 m_1, m_2 称为 Choice_i^w 等价, 直觉上, 意味着 i 在 w 处的将来选择有两种等可能结果即 m_1, m_2 处的历史集 H_{m_1}, H_{m_2} 。

Stit 框架 (Stit-Frame): Stit 理论的框架定义为有序组

$F = \langle \text{Tree}, \leq, \text{Agent}, \text{Choice}, \text{Instant} \rangle$, $\text{Choice} = \{ \text{Choice}_i^m : m \in W, i \in \text{Agent} \}$, 一个 Stit-模型于是可定义为 $M = \langle F, v \rangle$, v 是赋值函数, 它把每一个命题常元与 (m, h) 联系起来, 并给出真值集合 $\{1, 0\}$ 中的一个值。我们可以刻画 Stit_iP 的能动性: 即它们之间的可满足关系, 为方便计, 把 $M, (m, h) \models$ 记为 $(m, h) \models$, 于是:

(12) $(m, h) \models \text{Stit}_i P$ 当且仅当存在 $w \in W, w < m$ 使得:

(i) 对于所有与 m Choice_i^w - 等价的 $m', (m' \in W)$, 有: $(m', h') \models P$ (对于所有 $h' \in H_{m'}$)

(ii) 也存在 $m'' \in \text{instant}(m \in \text{instant})$, 使得 $w < m''$, 且 $(m'', h'') \not\models P$, 这里 $h'' \in H_{m''}$ 。

比较一下 Von Wright 与 Belnap 的能动性是很有意义的。Belnap 认为一个行动并不一定导致一个确定的结果状态, 这与 Von Wright 的主动性不同: 行动的作用在于它使选择进入一个确定的状态成为可能。可以看出, 这两者的差别在哲学上的对应于决定论与非决定论的

差别。

对比一人 Von Wright 与 Belnap 的模型定义 知道其满足关系都是在单个时刻来定义的。这引起了所谓的动作时间的界限问题 (Thomson,1977)。就我们的直觉来说,当我们说个别的动作在某个瞬间发生,这只是一个方便的(有时引起误解的)说法,因为我们都知道,一个动作的产生也有相应的两个相继的持续阶段。这些模型的形式定义都是在 (m, h) 处定义,这必须假设动作也必须在瞬间完成,开始与结束都必须在精确的瞬间发生。而这并不是得到保证的。那么在此时评价一个事态的真与假原则上是不可能的。

另外一个问题:如何解释引起 (Bringing it about that P) 句 BP 句子?例如:

$$BP \rightarrow P$$

Von Wright 解读为: $BP \rightarrow P$ 就是就在某个时刻 a 引起 p 发生,而 p 已经在那个时刻发生了。而行动产生这前, P 并没有被“引起”而发生[Von Wright,1983]。这显然是反直觉的。虽然有些哲学家建议将动作定义为时刻的有序对 $(m_1, m_2), m_1, m_2 \in W$ 是可能世界中的两个点。但这引起了另外一个问题;两个行动 $a_1 = m_1, m_2, a_2 = (m_1, m_2)$, 但 a_1, a_2 是直觉上看起来不同的行动,只是由于发生—结束的瞬间一致,怎么区别这两个行动?依有序对定义的, $a_1 = a_2$, 这有悖于直觉。

Belnap 的 Stit 理论分析也面临同样的难题。Choice_i^w 等价概念是在初始状态定义的,因为 m 是选择的起点。如果 agent_i 在 m 处引起 P , 即保证 m 处引起 P 为真,但 m 处的真值与早先时刻 $w < m$ 处真值有关,而 w 处的动作必须到 m 处,这样选择是在一个更早的时刻就开始了,这同样不符合直觉[L. ?qvist, 1974]类似 $BP \rightarrow P$ 的困难再次出现:因为按照 Belnap, $Stit_i P \rightarrow P$ 是公理,显然,它遭遇了与 Von Wright 同样的反直觉困难。不过, Belnap 的出了另外一种真理的行动性理论分析,即慎思行动性 (deliberative agency), 而区分上面 (12) 提到的完成行动性 (achievement agency)。慎思行动性由下式定义[Horty, Belnap, 1995]

$$(13) (m, h) \models D - Stit_i p, \text{ 当且仅当}$$

$$(i) \text{ 对于所有 } h' \in \text{choice}_i^m(h), \text{ 有 } (m, h') \models p;$$

$$(ii) \text{ 存在 } h'' \in H_m, \text{ 有 } (m, h'') \not\models P。$$

这个定义避免了选择的时不定问题：在 m 处的选择不会依赖于小于 m 的时刻。但这引起了另外一个问题；在 m 处的 agent 行动性既然是保证 agent 在 m 处使 P 为真， $(m, h') \models P$ 对 $h' \in \text{choice}_i^m(h)$ 特别地 $h \in \text{choice}_i^m(h)$ ，显然有 $(m, h) \models p$ ，在同一时刻，行动并未展开，何来保证 m 处 p 一定为真？或者说，在 m 处保证为 P 真，并不等于在起点时刻 P 就为真。许多哲学家都不约而同地指出了这一点[K. Segerberg, 1992; R. Hilpenen, 1997]，他们都不同意 D. B. Callasc 对此点的辩护[B, Challas, 1969]。

但是关于 Stit 理论有一点是忽视了，在有些情况下， $\text{Stit}_i P$ 的定义的条件是很难满足的。以下列树状框架来说明。

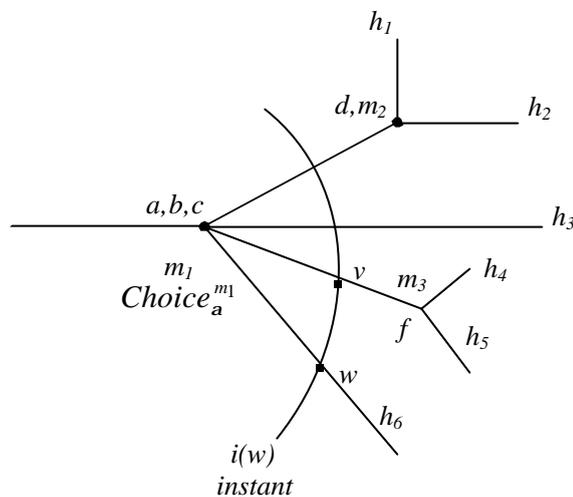


图 1 agent a 的选择图

依定义， $\text{Choice}_i^{m_1} = \{\{h_1, h_2\}, \{h_3\}, \{h_4, h_5, h_6\}\}$ ，agent i 在 m_1 处在三个动作 a, b, c 因而相应的三种选择。 $\text{Choice}_a^{m_2} = \{\{h_1\}, \{h_2\}\}$ ，而 $\text{Choice}_a^{m_3} = \{f\}$ ，(因为它有两个动作 d, e) 即一个动作不能产生两种选择，因为 Belnap 认为每个动作对应可一个历史。这个困难可以通过下列方式化解：此时动作 f 对应于一个集合 $\{h_4, h_5\}$ ，但意义并不那么明确，这是被 Belnap 所承认的那种情形。另外一处情形也是较为重要的，是如图 1 所示， $v, w \in \text{instant}$ ，不妨记为 $i(w)$ ，即同 v, w 一瞬间属于 instant， $m_1 < v, m_2 < w$ ，假设 $v \text{ choice}_i^{m_1}$ - 等价于 w ，那么必有 w 在 v 的选择的某历史上， v 必在 w 的选择的某历史上，于是有 $v \leq w$ 和 $w \leq v \Rightarrow v = w$ ，虽然 v 与 w 是关于某瞬间等价，但 $v \neq w$ ，因此， $\text{choice}_a^{m_1}$ - 等价概念的要求在 Belnap 的定义的条件集中是很难满足的，因而定义的存在性不能得到充分保证。这种冲突的根源在

于把可能世界与瞬间概念混淆起来。

上述对行动性的理性都有这样和那样的合理与不合理的方面。下面我们就存在的问题提供另外一种解决思路，哲学上它是基于意图与 fluent 逻辑之上的新方案，借此。避免上述提到的诸多困难。

基于计划的意图与行动性

在这部分里我们讨论选择什么样的方案。一个值得推荐的方案是回到行动性的意图解释上去。由于涉及到意图，关于知识的表征与推理的逻辑或动态逻辑作为讨论的工具自然是引进来了。虽然像 Belnap 那样，不考虑行动性中的心理成份，但采用分割来表示行动性的成份——选择函数时，所引起的困难是不容忽略的。实际上，引用意识成份来解释行动性时，好处是很明显的：它使行动性的说明变得自然，在哲学上这点是很重要的[Wooldridge, M, R, Jennings, 1995]。在本文解决方案中，我们采用 Searle 的观点：因果性是在行动本身的过程中体现出来的，表现为随时随地的意图控制 agent 自身的动作：Searle 称此为行动中的意图 (intention-in-action) [Searle, 1983a]，以区别 Davidson 的事先意图 (prior intention) 所产生的行动 (primitive action) [D. Davidson, 1980a]。事先的意图是在行动开始之前就有的，它确保 agent 在 w 处 P 为真，这正是前面讨论过的行动性与因果性，它适用于慎思型 agent (deliberate agent)。行动中意图是在行动中产生的，它有一个适应方向：即从心智适应世界，并且随时调节 agent 本身的规划过程 (Plan)，其因果性是指向 agent 本身的。如此看来，一个行动有两个成份：心智成份，这就是行动中的意图；另外一个是由意图引导 agent 的动作或物理运动 (bodily movement) [O' Shaughnessy, 1991]，必须指出的是，本文不考虑无意图的行动，因为任何一个行动必须在是有意图的，无意图只是局限于某些特定的环境描述[D. Davidson, 1980b]。

为了表达行动中的意图，我们需要一个基本动作 $sense_f$ ，其中 f 是 fluent 公式， $sense_f$ 表示 agent 在行动或规划过程中，在某个世界或情境 w 中所执行的动作：见证或检查 f 在 w 中是否为真为假[D.Harel, 1984]，而且 $sense_f$ 动作还表达在动态规划过程中，agent 的认知状态的更新[B. Van Linder, Van der Hoek, J. Ch. Meyer, 1994]。对于意图来讲， $sense_f$ 则表明：agent 执行该动作以检查每个情境中的 f 是否为真，这体现了意图对 agent 动作的因果控制，如果一旦在某个情境中 f 为真，此时 agent 可以认为它的意图已经实现，即意指 fluent f 为真。而执行规划则是在因果性的范围进行。由行动中的意图或心智对世界适应的方向所限定。

下列概念是必须的 (a) 设 a 为一动作， h 为树框架中的历史， T 为时间集 (这采取为整数集)，其元素为 t_1, t_2, \dots ，且满足树框架的两个条件 (面向过去成链以及存在共同的过去条件)，换句话说， $W_{t_1} \leq W_{t_2}$ 当且仅当 $t_1 \leq t_2$ 。记 $[h, t_1, t_2]$ 为 agent 在历史 h 上 t_1 到 t_2 所执

行的动作的集合[P. Singh, 1995], $t_1 \leq t_2$ (即 $t_1 \leq t_2 \wedge t_1 \neq t_2$), 是动作执行的最大时间区间的两个端点, 之所以这样是为了避免时间的界限不定困难。若 $\mathbf{a} \in [h, t_1, t_2]$, 则 agent 在 h 上的 t_1 到 t_2 时间段发生。(b) 动作与状态转移函数 do_i . 首先是动作定义: Skip 是动作(空动作); $sense_f$ 是动作, $bm(i) \in BM$ 是动作, 若 $\mathbf{a}_1, \mathbf{a}_2$ 是动作, 则 $\mathbf{a}_1; \mathbf{a}_2$ 也是动作, 这里我们不考虑其它复合动作, 原因同 Cao Son, C. Baral 一样[Cao, Son C. Baral, 2001]。于是我们可以说 $((\mathbf{a}_1; \mathbf{a}_2); \mathbf{a}_3); \dots; \mathbf{a}_n$ 是有意义的, 它表示 n 个动作相继执行。状态转移数 do_i 定义如下: $do_i(\mathbf{a}, w)$ 是指 $agent_i$ 在执行动作 \mathbf{a} 之后所到达的状态或世界或情景。如果我们考虑到 T 与 W 之间的同构, 这样我们可以对于每一个 $t \in T$, 可以有一个 $w_t \in W$ 与之对应。若 $w' = do_i(\mathbf{a}, w)$, 则 $w \leq w'$, 若 $\mathbf{a} \neq skip$, 则 $w < w'$. 记规划 $Plan = (((\mathbf{a}_1; \mathbf{a}_2); \mathbf{a}_3); \dots; \mathbf{a}_n) = (\mathbf{a}_1, \mathbf{a}_2 \dots, \mathbf{a}_n)$, 则 $do_i(plan, w_t) \equiv do_i(\mathbf{a}_n, do(\mathbf{a}_{n-1}, \dots, do(\mathbf{a}_1, w_t))) \equiv do_i(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ 表示 $agent_i$ 在执行规划 Plan 的同时, 也经历了 w_t, w_{t+1}, \dots, w 等情景。(3) 我们修改 Belnap 的树状框架, 使模型适合于目前的具有动态规划的行动性。框架表示为 $M = \langle W, T, \leq, Intension, DO(i), \mathbf{p} \rangle$, 式中 W, T, \leq 定义如前, $Intension = \{[h, t_1, t_2]: h \text{ 是树中历史}, t_1, t_2 \in T\}$, $DO(i) = \{do_i(\mathbf{a}, w_t): \mathbf{a} \text{ 是任一动作集合}, w_t \in W\}$, $\mathbf{p} = \{Plan: Plan = (\mathbf{a}_1, \mathbf{a}_2 \dots, \mathbf{a}_n), \mathbf{a}_1, \dots, \mathbf{a}_n \text{ 是 } A \text{ 中任意动作}\}$ 。(4) 引进描述行动性的概念, 它表达 $agent_i$ 的意图指引的因果性, 即 $agent_i$ 动作规划的原因, 它最终指向某状态或情景, 在该情景中, $agent_i$ 意图某个 fluent 公式 f 为真。但在这过程中, 意图不是事先意图, 而是体现在动作规划的每一个阶段, $sense_f$ 动作体现行动中的意图: 即世界对心智的适应性, 它期望 $fluent f$ 在某个情景中为真, 通过这个动作 $agent_i$ 到达结果状态或情景, 因果性即意图对行动的保证与指引[Searle, J, 1983b]。因此, $agent_i$ 的行动性公式以表达为 $Guide \langle Intend_i f, \mathbf{a} \rangle, \mathbf{a} \in \mathbf{p}$, 它读为 $agent_i$ 意指 fluent 公式 f , 而动作系列或规划 \mathbf{a} 受意图指引。

关于行动性的模型可以定义如下:

$M, h, w_i \models \text{Guide}(\text{Intend}_i f, \mathbf{a})$ 当且仅当 :

(1) $\mathbf{a} \in BM$, 即 \mathbf{a} 是 agent 的身体动作 ;

(2) $\forall h' \in H(w_i)$ 和 $t_1, t_2, t_1 < t_2, \mathbf{a} \in BM, \mathbf{a}_i \in [h', t_1, t_2], t_1, t_2 \in T$, 有 :

当 $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2 \dots, \mathbf{a}_n) \in \mathbf{p}$ 时 , 存在状态转移出数 do_i 且使得。

(3) $do_i((\mathbf{a}_1, \text{sense}_f), (\mathbf{a}_2, \text{sense}_f); \dots, (\mathbf{a}_n, \text{sense}_f)) = w_{t_2}$ 且

(4) $h, w_{t_2} \models p$, 但 $h, w_{t_1} \not\models p$

(5) 对于任意 $1 < k < n$, 则 $h, w_{t_k} \not\models p$

(1) 要求 *agmti* 只对自己的身体动作负责 , 行动的意图指引涉及自我指涉及的因果性 , 正如 *Searle* 所分析的那样。(2) 要求这种指引是在某个历史中进行的 , 而是从现在时间 t_1 指向未来 t_2 , 即意图涉及未来。如果此时有一个动作规划 \mathbf{a} , 它使得 *agenti* 从 w_{t_1} 过渡到 w_{t_2} , 且在 w_{t_2} 处 P 为真。(5) 则陈述 *agmti* 在达到 w_{t_2} 之前 , P 在 w_{t_1} 与 w_{t_2} 之间的情景 w_{t_k} 中是为假的 , 此外 , 还需说明的是 ,(3) 使用 sense_f 动作 , 我们把它视为身体动作的一部分 , 这个直觉是很简单的 : 即 *agenti* 利用它的感官 (人类利用眼睛 , 机器人利用传感器) , 这些都可以看作是 *agenti* 的动作领域。而且我们要求 *agenti* 在每个动作 $\mathbf{a}_i (i = 1, 2, \dots, n)$ 之后马上进行检查即执行 sense_f 动作 , 这是行动中意图这个概念本身所要求的 : 因为 sense_f 动作随时都检查 fluent 的可满足状况 , 这一点与 Belnap 的 *Stitp* 的行动性 (事先意图) 是不一样的。

事实上 , 这个关于行动性的定义与前面诸行动性是有许多共同点的 , 例如都强调意图 (事先意图和行动中意图) 在因果性中的重要性 ; 都承认非决定论 ; 都承认行动性存在两个不同的成份 , 都承认动作使得情景变化和使得 agent 选择其某些行动路径 ; 都使用可能世界的框架。但不同之外也是很明显的 : 例如我们的行动性理论更突出意图对于因果性的支配地位 ; 强调规划对于行动性的现实意义 , 而不是用历史来刻画选择 , 而这并不一定导致意图的实现 , 从而行动性的实现是有疑问的。特别是引入 sense_f 动作 , 使得规划服从于将来意图的指导 , 它比事先意图更好地抓住了意图与因果性的关系。

从某种程度上看,事先意图是行动中意图的特例;因为事先意图一旦形成并且要实现的话,那么在规划进程中,不需要特别的检查行动 *sense_f* 来保证意图的实现,因此,规划的定义中 *sense_f* 这个动作可以略去,此时,

$do(\mathbf{a}_1; sense_f), (\mathbf{a}_2; sense_f), \dots, (\mathbf{a}_n; sense_f) = do(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ 而 $(\mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_n)$ 的存在

正是事先意图保证的。

本文是关于行动性理论的认真尝试。与其他行动性理论不同,它用 *sense_f* 动作和规划来表达因果性。在形式上,它采用了动态逻辑和 *flnent* 理论的概念,它与通常采用的可能世界既相似而又不相同:它用动态规划来代替可能世界之间的可达关系,就语义模型而言,它是 Stit 框架的扩充。更重要的是,它避免了前面总结的有关行动性理论的不足之处:动作可以看作是某历史上两个不同时刻的 agent 身体运动,且避免了时间界限引起的动作界定不利之处,也避免了类似 $BP \rightarrow P$ 有悖于直觉的语义有效公式出现。另外,它强调规划对于行动性的重要地位,这与局限于树状框架本身来谈选择而忽视其心智机制的 stit 理论既是补充又是深化。这种基于规划和 *sense_f* 行动的主动性方案,另一种值得探讨的竞争性方案。对于该理论方向所涉及的问题,例如其逻辑系统的公理化,逻辑表达式之间的异同将是值得进一步深入研究的课题。

注释

1、行动性对应的英文原词“Agency, 这样翻译是考虑到在行动哲学中, Agency 经常与 Agent 的某种促使行动性产生或阻止某行动发生联系在一起,如果对应于 Agent, 翻译成“主体性”,这与其他哲学领域的主体性不同。徐明先生将 Agency 翻译成“主动性”,似乎是不妥的,因为 Agency 有被动性引起的一面,况且“主动性”在行动哲学中对应于“activity”。关于徐明先生的翻译,见文“分枝时间上谈因果(I): 过渡,事件与原因”,载于《当代基础理论研究卷——留美哲学博士文选》,牟博编,商务印书馆,2002, P237 页附注,另外本文对 Agent 一词不予翻译,直接使用英文称呼拼写。

2、关于 Belnap 对 Agent 的定义,可参见 Belnap 于 2002 年 4 月在中山大学逻辑与认知研究所的讲演稿的附录,该讲演题为“Agents in Dranching Space-times”。

3、在 Von Wright 的分析里, World 与 Situation 是同义上使用的,这一点作者在本文后面也是这样使用的。有关情况可见[Hilpenn, R, 1997]

4、事实上， $e(u),d(u)$ 应如下定义：

$$e : W \times A \rightarrow W$$

$$d : W \times A \rightarrow W$$

这里 A 为所有原子行动的集合，若 $a \in A$ 固定，则 $e(u, a), d(u, a)$ 可简写为 $e(u), d(u)$ 。

而在本文中， a 不能略去的，它对本文所提出的行动性理论有密切关系。

5、相关内容可在注释 2 所提到的 Belnap 的演讲稿中找到，在该讲演中，Belnap 报告了 stit 理论的哲学逻辑上的总结及相关研究内容与进展。

参考文献

- [1] ?qvist, L, A new approach to the logical Theory of Actions and causality,[A] in Stenlund, S (ed), Logical Theory and Semantic Analysis[C], Reidel, pp.73-91, 1974.
- [2] Belnap, D., Backwards and forwards in the modal Logic of agency [J], philosophy and phenomenological research, 61, pp777-807, 1991.
- [3] Belnap, N, Perloff, M, Xu, M. Facing the future: agents and choices in our indeterminist world[M], Oxford University Press, 2001.
- [4] Cao Son, Baral, C., Formalizing sensing actions: -a transition function based approach[J], Artificial Intelligence, pp19-91, 2001.
- [5] Chellas, B, The Logical form of imperatives (Dissertation, Stanford University)[D], Perry Lane Press, p66, 1969.
- [6] Davidson, D., Agency[A], in essay in actions and events[C], Davidson, D (ed,) Clarendon. Press, pp43-61. 1980a.
- [7] Davidson, D., Essays on actions and events[M], Ch.3. 1980b.
- [8] Harel, D., Dynamic logic[A], in M. Gabbay, F, Guentner(eds), Handbook of philosophical logic[M], vol 2, Ch. 10, Reidel. 1984.
- [9] Hilpinen, R., On action and agency[A], in Logic, action and cognition[M], E. Emerged, S. Lindstorm (eds), pp3-5 , 1997.
- [10] Hornsby, J., Agency and causal explanation[A], in the Philosophy of action[C], R. Mele(ed). Oxford University Press, 1993.
- [11] Kanger, S Law and Logic[J], Theoria, pp105-281. , 1972.
- [12] Horty, F, N, Belnap, The deliberative stit: A study of action, omission, ability, and obligation[J]. Journal of philosophical Logic, Vol 24, No. 6, pp583-644, 1995.

- [13] Mele, R. (ed). The Philosophy of action[C], Oxford University Press. pp1-2. , 1997.
- [14] Lewis, D. Counterfactuals[M], Basil Blackwell, Oxford , 1973.
- [15] Lewis, D., Postscripts to “causation”[A], in D. Lewis, Philosophical papers () [C], Oxford University, pp172-213. 1986.
- [16] O'Shaughnessy. Searle's theory of action[A], In E. Lepore, Van Gulick(eds.), John Searle and his critics[C], Basil Blackwell. pp271-287 1991
- [17] Pörn, I, Action theory and social science[M]; Reidel, p7, 1997.
- [18] Searl, J, Internationality: An essay in the philosophy of mind[M], Cambridge University Press, pp83-98, 1983.
- [19] Searle, J, 同[18] p93.
- [20] Singh, P.. Multiagent systems, A theoretical framework for intentions, know-How, and communications[M], Springer-Verlag. p21, 1995.
- [21] Thomson, J., The time of a killing[J], the Journal of philosophy, 68, pp115-132. ,1971.
- [22] Van der Hoek, Van Linder, Ch.-Meyer, A logic of capabilities[A], in Proceedings of the third international symposium on the logic foundation of computer science (LFCS'94)[C], A. Nerode, V. Matiyasevich(eds), pp366-378, Springer-Verlag, 1994.
- [23] Von Wright, Norm and action[M], Rutledge & Kegan Paul, 1963.
- [24] Von Wright, Norms, truth and logic[A], in his Practical reason: philosophical paper (I)[C], Cornell University Press, pp130-209, 1983.
- [25] Woodridge, M, Jennings, N, Intelligent agents: theory and practice[J], The knowledge engineering review, 10(2), pp115-152, 1995.

Agency, Causality and Intention

FEI Ding-zhou.

(Department of philosophy, Wuhan University, wuhan, 430072)

Abstract: How to describe agency is the center topic. It is the tasks of many studies to seek to give philosophical analyses and logical expressions for agency. This paper includes a few of main theories about agency, which is closely related with causality, and shows the their own semantics, mostly in possible world frame respectively. However, there are some problems induced by them, so a new treatment for previous theories will advance discussion of agency, if we adopt the Intention-in-action from Searle's notations in philosophical aspect, at same time, dynamic logic with plan as logic tools. Finally, this paper compares my investigation with others. and suggests some further issues.

Key words: Agency, Tree-like frame, prior intention, stit theory, Intention-in-action.