doi:10.13582/j.cnki.1672-7835.2024.01.005

信念-愿望-意图逻辑的哲学基础

郝一江
(中国社会科学院 哲学所,北京 100732)

摘 要:从常识性心理学、心灵哲学和行为哲学的视角,系统探讨意图、信念、愿望、规划与 Agent 实践推理的关系,可得到如下主要观点:(1)意图既是潜在的行为影响者,又是行为控制者,它为进一步的实践推理和规划提供了重要的输入。(2)意图涉及承诺;规划揭示了"以实践推理为中心的承诺的重要属性";资源受限 Agent 的规划通常是具有层级结构的局部规划;意图是规划和实践推理得以关联的关键因素。(3)规划支持协调,并能够扩大慎思对之后行为的影响。(4)意图不仅能够帮助实践推理确定选项的相关性和可接受性,而且还能够构建"权衡愿望-信念理由"的过程。

关键词:信念;愿望;意图;规划;Agent 实践推理
中图分类号:B81 **文献标志码**:A **文章编号**:1672-7835(2024)01-0033-08

一 问题的提出

随着人工智能时代的到来,具有理性行为能力的计算 Agent(主体或智能体)的研究受到了广泛关注。信念和意图的概念在自主 Agent 的设计与实现中起着核心作用。这些概念不是源自人工智能和多 Agent 系统,而是来自心灵哲学。心灵哲学认为它们是 Agent 的基本心智态度:信念有一个"从心智到世界"方向的适应过程,即 Agent 试图调整信念以适应世界真相;而意图有一个"从世界到心智"方向的适应过程,即 Agent 试图使世界与他们的意图相匹配。

当我们认真思考"面向未来的意图和规划以及它们对下一步的实践推理的影响时",我们的心智状态和理性 Agent 会发生怎样的变化呢?Bratman(1987)[①] 的专著《Intention, Plans, and Practical Reason(意图、规划与实践推理)》主要探讨了这一问题。这部专著不仅给出了信念、愿望和意图这些基本概念,而且说明了它们之间的关系,这为之后在人工智能领域得到广泛应用的信念-愿望-意图逻辑[②]奠定了很好的哲学基础,因而受到诸多学者的推崇和引用,截至 2023 年 11 月 18 日,这部专著被引用 5 049 次。

Bratman 仅仅给出了信念-愿望-意图理论的半形式化系统,Cohen 和 Levesque[③] 以及 Rao 和 Georgeff[④] 在 Bratman 的基础上,从不同的视角给出了不同的形式化系统,并且由于其极高的引用率,分别于 2006 年和 2007 年获得了自主 Agent 和多 Agent 系统的影响文章奖。本文将在 Anscombe[⑤]、Goldman[⑥]、Davidson[⑦] 和 Bratman[⑧] 等的相关研究成果基础上,结合现代人工智能中的 A-

收稿日期:2023-10-08
基金项目:国家社会科学基金后期资助项目(22FZXB092);中国社会科学院哲学研究所创新工程项目(2024ZXSCX06)
作者简介:郝一江(1971—),男,山西阳泉人,博士,副研究员,主要从事科学哲学和现代逻辑研究。

①Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987.

②张晓君:《信念-愿望-意图逻辑及其应用研究》,中国社会科学出版社 2017 年版。

③Cohen P R, Levesque H J. "Intention is Choice with Commitment", *Artificial Intelligence*, 1990, 42(3): 213-261.

④Rao A S, Georgeff M P. "Modelling Rational Agents within a BDI-architecture", Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, 1991, pp. 473-484.

⑤Anscombe G E M. *Intention*. Ithaca: Cornell University Press, 1963.

⑥Goldman A. *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall, 1970.

⑦Davidson D. *Essays on Actions and Events*. New York: Oxford University Press, 1980.

⑧Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987.

gent 理论[①],重点探讨信念、愿望、意图、承诺等对 Agent 规划和行为的影响,从而为信念-愿望-意图逻辑奠定坚实的哲学基础。

目前,有大量已实现的信念-愿望-意图(Belief- Desire-Intention,简称 BDI)逻辑系统,这些系统支持创建使用信念、愿望和意图这三种结构定义的 Agent;而且还存在对这些功能进行扩展的系统,以便提供额外的人类推理能力。所有这些已实现的 BDI 系统的共同点是:都使用了规划库的概念。每个规划库由以下三部分组成:规划组件、(规划能够实现的)目标或意图、(规划适用的)背景。规划组件由以下两部分组成:(直接影响环境的)行为、(在需要时扩展到其他规划中的)子目标或意图。规划中的目标嵌套允许构建规划-目标层级结构,其中目标节点的分支指向"能够实现此目标的规划",规划节点的分支指向"完成此规划必须实现的目标"。已实现的 BDI 系统还包含一个 BDI 执行引擎,该引擎指导 Agent 的推理过程。不同引擎使用的算法可能存在差异,但是基本版本的 BDI 执行引擎主要轮廓包括:感知、更新、意图修正、规划筛选、方案选择、行动[②]。

我们对自己和他人的许多理解都植根于一种常识性的心理框架,即一个以意图为中心的框架。在这个框架内,使用意图的概念来描述人们的心智和行为,即利用意图表征人的心智状态。意图的重要性体现在:(1)它们与广泛的情感反应、道德态度和法律制度紧密相连;(2)意图为我们每天预测他人做什么、解释他人做了什么以及协调我们的规划与他人的规划提供了基础。

做知识级分析的系统被称为 Agent[③],它是人类智能、动物智能和机器智能的统一模型[④],可定义为从感知序列到实体动作的映射[⑤]。全文着重探讨了信念、愿望、意图、规划与 Agent 实践推理的关系,以解释如何从意图的角度来表征 Agent 的心智和行为。为此,需要探讨规划、意图、行为和实践理性的关系[⑥]。

二 规划

人是制定规划的 Agent,这些规划支配我们以后的行为。为此,需要两个核心能力:(1)具有"有目的地采取行动"的能力;(2)具有制定和执行规划的能力。对未来规划的需求植根于如下两个普遍的需求[⑦]:

第一,需要我们是理性的 Agent。这在一定程度上意味着慎思(deliberation),更一般地说,理性反思有助于塑造我们的行为。成功慎思的程度显然是受到时间和其他资源的限制。因此,允许慎思和理性反思超越当下,以影响行为。

第二,需要 Agent 协调。为了实现复杂的目标,首先,Agent 必须协调他现在和未来的活动(即个人协调);其次,Agent 还需要协调他的活动与别的 Agent 的活动(人际协调)。通常同时需要这两种协调。因为我们人类既受时间、金钱等资源的限制,也是社会的 Agent;换句话说,人类是资源受限 Agent。

因此,我们需要找到慎思和"能够超越当下的理性反思"影响行为的方法,而且还需要找到来自个人协调和人际协调的支持资源。通过为未来制定更大的规划来促进协调,这些规划有助于随着时间的推移,协调我们自己的活动以及我们的活动与其他人的活动。现在制定以后的规划,就能使得现在的慎思影响以后的行为;即把慎思的影响扩展到现在之后。

我们通常只确定局部规划,然后根据需要和时间的推移逐步填充这些规划。Agent 规划的这种不完全性创造了对推理的需求,这种推理可以刻画制定规划的 Agent,因为推理以给定的初始的局部规划为基础,旨在用适当手段、初始步骤或者相对更具体的行动方案的流程来填充这些规划。

①蔡自兴,徐光佑:《人工智能及其应用(第 4 版)》,清华大学出版社 2010 年版,第 311—339 页。

②Adam C, Gaudon B. "BDI Agents in Social Simulation: A Survey", *The Knowledge Engineering Review*, 2016,31(3): 207-238.

③林颖,张晓君:《信念-愿望-意图逻辑探析》,《重庆理工大学学报(社会科学版)》2016 年第 3 期。

④陈亚楠,郝一江:《基于偏好排斥等级 BDI 主体的决策行为研究》,《重庆理工大学学报(社会科学版)》2022 年第 9 期。

⑤张晓君,邱君:《基于改编命题动态逻辑的 Agent 交互协议推理》,《湖南科技大学学报(社会科学版)》2022 年第 5 期。

⑥Bratman M E. "Intention, Belief, Instrumental Rationality", Sobel D, Wall S. (eds.), *Reasons for Action*. Cambridge: Cambridge University Press, 2009, pp.13-36.

⑦Bratman M E. *Intention, Plans, Practical Reason*. Cambridge: Harvard University Press, 1987, pp.18-19.

三　意图

常识性心理学承认意图是一种心智状态,并且允许我们把行为描述为意图做某事或者有特定的意图。因此可以从正在意图的行动的状态开始讨论意图。首先说明什么是“意图做某事”,然后解释“意图做某事或者有特定意图”,这需要借助于“有做某事的意图”的概念。

当我们谈论“有做某事的意图”时,经常会想到关于未来的意图。面向当前的意图应该引导我们把“面向未来的意图”视为核心意图。要理解“正在意图做某事”的概念,就必须理解“面向未来的意图”的概念①。

行为主义的支持者认为:面向未来的意图嵌入在当前的意图活动中;对未来行动的承诺总是嵌入在当前行动中。但是这过于简单化了,因为只有一小部分未来行为在我们的意图范围之内,现在承诺的只是未来行为范围的一小部分。因此,对未来行动的承诺似乎不仅仅隐含在目前的意图活动中。但是,除了目前的意图活动之外,我们对未来行动的承诺还包括什么呢?

当我们试图回答这个问题时,很快就会面临一个令人费解的三重困境。我们今天的意图并没有提升到可以神灵般地进行时空变换,并控制明天的行动。面向未来的意图大概不是不可改变的。如果这种意图是不可撤销的,这显然是不合常理的;毕竟,事情在变化,我们并不总是能够完全准确地预测未来。这意味着,除非某个时候有理由重新制定规划,否则为什么现在要费心决定未来做什么呢? 因此,面向未来的意图似乎:(1)在形而上学上是令人反感的(因为它们涉及较远未来的行动);(2)或者在理性上是令人反感的(因为它们是不可撤销的);(3)或者只是在浪费时间。

这三重困境导致了对“面向未来的意图”概念的怀疑,这种怀疑一旦产生,就会迅速蔓延到“正在意图的行动”的概念,因为“正在意图的行动”的核心就是面向未来的意图。如果面向未来的意图导致我们需要在较远的未来采取行动,或者把此类意图看作是不可撤销的意图或者浪费时间的意图,那么我们应该对“正在意图的行动”的观点持怀疑态度。

为了摆脱这些困惑,Bratman(1987)提出了意图的替代方法,这种方法已经成为心灵哲学和行动哲学的主要传统。这一方法成功地描述了关于意图的一般方法,认为:我们不应该从正在意图的行动的状态开始讨论,而应该直接对行动中出现的意图进行讨论,即直接对“意图做某事或者有特定意图”进行讨论。意图的替代方法具有如下四个论点:

(1)行动中的意图具有方法论上的优先性。当我们遵循这一策略时,自然会产生这样的想法:一个行为是意图进行的行为,或者是有某种意图的行为,是因为该行为与 Agent 的愿望和信念之间具有事实上的关联。

(2)关于行动的意图-愿望-信念理论。根据 Agent 的愿望和信念以及“与这些愿望和信念具有因果关系的”行为,来理解正在意图的行动和已经完成的意图行动。因此,行动意图的常识概念的基本结构涉及两种基本的心智状态:一是那些具有信念作用的心智状态;二是那些具有愿望作用的心智状态。正是由于与这些心智状态建立了因果关系,才使得某个行动有意图,或者是以某种意图去做。

(3)具有推广策略。如果我们接受了“意图在行动中的优先性”和这里的意图-愿望-信念理论,那么就会很自然地把对行动意图的描述,直接推广到面向未来的意图的描述。换句话说,一旦我们有了一个关于“有意图的行动和带有某种意图的行动的”充分描述,就可以期望获得所有的主要材料,进而就可以处理面向未来的意图。

(4)可以把面向未来的意图还原为适当的愿望和信念。从 19 世纪哲学家奥斯汀(Austin)的著作中可以找到这种简化方法的经典陈述:对 Agent 心智和行动的常识性描述只需要两个参数,即愿望-信念框架。

这里的观点(1)和(3)表达了研究策略,而(2)和(4)是关于 Agent 心智和行动的实质性观点,因此(2)和(4)一起构成了意图和意图行动的愿望-信念模型,这是当代心灵哲学和行为哲学的主流观点。

但是,Bratman 认为,应该拒绝这四个论点和

①Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.20.

愿望-信念模型①。这是因为:我们对意图的理解,在很大程度上依赖于对面向未来的意图的理解。人类是从事规划的 Agent,而面向未来的意图是形成更大规划的一部分,这些规划在协调以及之后进行的实践推理中发挥着独特的作用,从而使我们能够通过当前慎思来影响未来的规划。换句话说,意图是这些规划的基石;而规划是扩大了的意图。

关于替代意图方法的四个论点更适合于理解"不会制定规划的"Agent(例如,狗和猫)。由于人类是能够制定规划的 Agent,因此这四个论点并不适用于人类。换句话说,这四个论点没有正确理解"人作为制定规划的 Agent 的概念"。

四　意图与规划

Bratman 对意图的理解具有心灵哲学的功能主义传统。因为他对各种心智状态的常识性理解,依赖于对这些状态所嵌入的适当的、潜在的规则的假设。这些规则把这些不同的状态联系在一起,伴随着相关的心理过程和活动,具有其独特的"输入"和"输出"——感知和行动。虽然这些规则有时可能涉及某种方式的行为倾向,但是也可能涉及以独特方式进行思考或者推理(或者抵制推理)的倾向②。

(一)规划理论

常识心理学的许多核心规则涉及相关的规范或者标准。例如,当认识到我们持有的信念总体上是不一致的,通常倾向于朝一致性的方向修正信念;与这种倾向性相关联的是规范,因为规范要求信念具有一致性。在对常识框架的规则网络进行解释时,需要在一定程度上说明这些规范或者标准是什么。

对意图的怀疑,特别是对"面向未来的意图"的怀疑,说明意图可以嵌入到这样的规则和规范的网络中。我们的常识框架把意图视为一种独特的心智态度,而不是把意图与普通的愿望和信念混为一谈,或者把意图简化为普通的愿望和信念。作为意图特征的规则和规范,包括资源受限理性 Agent 的持续实践推理和行动中的局部规划的角色特征。

意图规划理论需要探讨如下两个问题:(1)如何避免对意图的怀疑?(2)对于像我们人类这样的"在关注问题、选项的慎思、确定可能的后果、执行相关计算等方面资源受限"的 Agent 来说,应该如何理解实践理性的内涵?通过反思意图和局部规划"在资源受限理性 Agent 的不断发展的实践推理和规划中的作用",就可以很大程度上解决第一个问题。

(二)Buridan 实例

这种对意图规划的探讨,需要对"作为制定规划的 Agent"进行探讨,因为 Agent 的资源受限而且有协调需求,包括对社会协调和个人协调的需求。Agent 对"支持协调规划的需求"至少可以部分归因于"Agent 的资源受限"。但是,Agent 对协调的需求也会给"作为制定规划的 Agent"带来压力③。

例如,一头有理性的驴被放在'两堆具有同样吸引力的干草'中间,可能会饿死,因为它没有理由走向左边草堆而不是走向右边草堆,也没有理由走向右边草堆而不是走向左边草堆。这种"在 Agent 看来同样可取的选择之间作出选择的实例"被称为 Buridan 实例。

(三)普通意图

为了探讨意图规划理论,需要关注普通意图的简单实例,在这些实例中,面向未来的意图和局部规划不会因为"先前的慎思"给"之后的行为"带来困扰。如果把面向未来的意图作为意图的范例,就会对其产生扭曲的看法。为了避免这种情况的发生,需要从普通意图开始讨论。

五　通向规划理论的道路

现在探讨愿望-信念模型,并且描述意图的几个特征,这些特征是后文讨论的核心内容,因此,需要把意图视为独特的心智状态;之后将描述一些问题,这些问题将影响以后的讨论。

(一)愿望-信念模型

愿望-信念模型的基本观点是:如果一个行动与 Agent 的愿望和信念有着恰当的关系,那么这个行动就是意图行动,或者是带有某种意图的

①Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.24.

②Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.27.

③Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.29.

行动。在探讨意图和行动模型的过程中,会涉及关于实践理性的问题,即什么使得行动合理?Agent 在某一时间的愿望和信念,为他当时采取的各种行动提供了理由①。

(二)意图和承诺

由于意图涉及承诺,因此意图理论的主要工作之一,是对意图中涉及的承诺进行全面的描述。第一步是区分承诺的两个方面。承诺的第一个方面涉及意图和行动之间的关系,称之为承诺的意志维度(简称为意志承诺)。意图和愿望都是支持 Agent 行为的心智态度,但是普通的信念则不是。意图是行为控制者,而普通愿望则不是,因为普通愿望仅仅是行动的潜在影响者。面向未来的意图中的承诺的意志维度,来源于意图是行为控制者这一事实。承诺的第二个方面是面向未来的意图。至关重要的是面向未来的意图在其初始形成和最终执行之间所起的作用,这些作用构成了以推理为中心的承诺维度。可见,面向未来的意图的承诺特征有两个维度:意志型和推理型,二者密切相关②。承诺的意志和推理这两个维度的结合涉及 Agent 的协同,即把这两个承诺维度放在一起,有助于解释意图如何在支持 Agent 的个人协调和社会协调方面发挥其特有的作用。

(三)愿望-信念模型的适度扩展

事实上,愿望-信念模型面临着挑战:因为与承诺的每个方面相关的基本规律和倾向,及其对实践理性的规范性概念的影响,"在严格遵循愿望-信念框架的情况下"很难得到充分的描述。这是因为在给予意图作为独特的心智态度的地位时,对愿望-信念模型的描述还是不够充分。意图是独特的心智状态,在 Agent 思考和行动之间起着独特的因果作用。但是在发挥这些作用时,意图并没有提供"与它们所采取行动的合理性直接相关的"考虑因素。意图可以间接地与理性行为相关。对此,需要区分这种相关性可能发生的三种方式。首先,意图可能具有间接的实际意义。如果 Agent 的愿望与实现早期意图有关,就会发生这种情况。第二,意图可能具有间接的认知关联。Agent 可能会把他先前意图 A 视为将要完成 A 的证据,在进一步的推理中,把正在拥有的意图 A 视为理所当然的。第三,意图可能具有间接的二阶相关性。Agent 可能会把他先前对 A 的意图视为证据,证明正在拥有的意图 A 事实上受益于他目前的愿望-信念理由的平衡;或者从他目前的愿望和信念的角度来看,作为"重新考虑这一意图的代价是不值得的"的证据。事实上,先前意图与行为合理性的相关性至多是间接的,因为它是通过 Agent 的愿望和信念来实现的。换句话说,Agent 的愿望和信念提供了"与其行为合理性直接相关的"考虑因素。

这种观点被称为愿望-信念模型的适度扩展,因为它保留了该模型的实践理性概念;而且在对愿望-信念模型的描述性进行修改的同时,还保留了该模型的规范性。但是这种扩展具有一定的局限性,因为一旦仔细考查意图的作用,就会把意图作为进一步意图推理的输入,这时对实践理性的规范性解释就会变得复杂。

现在讨论面向未来的意图。假定意图是与愿望和信念一致的心智状态。这种心智状态有如下三种倾向:保留该意图而无需重新考虑的倾向;从保留该意图到进一步意图的推理倾向;根据该意图约束其他意图的倾向。这些倾向是行为控制的支持态度;而且这些倾向抵制反思,因此意图有一种特有的惯性。意图发挥着独特的作用,作为进一步实践推理的输入,以达到进一步的意图③。在每种情况下,都可探讨这些倾向对实践理性的规范概念的影响。

现在只关注"意图作为进一步实践推理的输入"的相关问题。例如:假设我在今年二月时打算五月份去成都看望儿子。在形成这一意图之后,我可能会继续从中推断出一个更具体的意图(例如,在五月的第一个星期内去),或者关于手段的进一步意图(例如,乘坐动车去),或者关于初始步骤的进一步意图(例如,请求我的同事承担我的教学任务)。

为了简单起见,把重点放在我对"一个关于手段的意图"进行推理的情况中。在这样的推理中,我开始打算去成都,并且考虑如何去。我去成都的意图与"我在这种推理中达成的'关于手段

①Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.35.

②Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.28.

③Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.47.

的进一步意图'的合理性"直接相关。例如,这与我决定乘坐动车的合理性直接相关。当我真正乘坐动车时,我看到如下这样的事实:行动是实现我的意图(去成都)的一种手段,与它的合理性直接相关。我先前打算去成都,这与我后来打算乘坐动车的合理性以及我最终采取的行动直接相关。但是,这种"对我先前意图的作用的理解"与"只有愿望-信念理由才具有这种直接相关性的观点"相冲突。

这种冲突在下面的 Buridan 实例中尤其明显,在这些实例中,我们可能形成了一种倾向:支持几个不相容但是(据我所知)同样可取的选项中的一个。例如:虽然我倾向于从北京南站出发走京台高速到南京南站的愿望-信念理由,似乎与走京沪高速的理由同样多,但是我还是必须作出决定。碰巧的是,我(可能是武断地)决定走京台高速。现在我必须弄清楚如何到达那里,在这种推理中,我打算到南四环中路右转经过德贤路到京台高速。

在这一手段-目的推理中,我到南京南站的意图与"我先前打算到南四环中路右转经过德贤路的衍生意图"的合理性直接相关。这就是为什么我们认为我右转的意图是理性的,而左转到榴乡路的意图是非理性的;但是在我形成走南四环中路意图之前,情况并非如此。既然决定走京台高速,我想我应该在南四环中路右转;但是在我决定走京台高速之前,我并不这么认为。因此,我先前打算走京台高速与我在南四环中路右转的合理性直接相关,这一分析与实践理性的愿望-信念模型相冲突。由此可见,Agent"对这种手段-目的推理的常识性理解"与"愿望-信念模型"之间存在着严重的冲突关系。如果认真考查这种常识性理解,就会发现愿望-信念模型是有问题的。

与实践理性的愿望-信念模型相关联的是实践推理模型,这种模型可以对各种选项的愿望-信念理由进行权衡①。这种权衡无疑是实践推理的一种常见形式。但是,一旦考虑到"意图是一种独特的心智态度",就必须接受关于达到预期目的的手段的推理。在这种推理中,我们把先前意图视为与"进一步意图和行动的合理性"具有某种直接关联性,即在愿望-信念模型中,保留了愿望-信念理由。为此,需要一个令人满意的理论来解释这两种推理是如何相互关联的。

(四)基于意图的推理

那么,应该如何理解"先前意图提供了衍生意图和行为的合理性直接相关的考虑因素"呢?一种方法是在愿望-信念模型提供的行动推理结构的基础上,用基于意图的推理补充对行动推理的解释。意图不仅仅是 Agent 独特的心智状态,而且还提供了行动的理由,这些理由是超越了普通愿望和信念的理由。正是通过提供这些行动理由,意图提供了"与手段-目的推理结论的合理性"直接相关的考虑因素。

六　规划与实践推理

关于未来行动的意图通常是制定更大规划的要素,这些规划有助于促进社会和我们自己生活的协调,还有助于使得先前的慎思能够塑造后来的行为。我们确定了先前的局部规划,只有在遇到问题时才会重新考虑这些规划。提前确定这些规划的能力,使得我们能够实现复杂目标。这种协调规划的能力是一种通用手段,它在追求各种不同的目标时有着重要的作用。

(一)作为心智状态的规划

首先需要区分作为抽象结构的规划和作为心智状态的规划。当我们谈论规划时,我们脑子里有一种特定的心智状态,而不仅仅是一种抽象的结构,比如说,可以用一些博弈论符号来表示。以下只讨论作为心智状态的规划,包括对行动的适当承诺:只有当我们的意图是真实存在时,我们才有一个规划。

这样理解的规划,其意图的作用是巨大的。它们都有意图的如下特点:意图抵制重新考虑,在这个意义上说意图具有惯性;而且意图不仅仅是潜在的行为影响者,而且还是行为控制者;意图为进一步的实践推理和规划提供了重要的输入。但是,与相对简单的意图相比,由于意图复杂性的增加,规划揭示了"以推理为中心的承诺的重要属性"。特别是,像我们人类这样的资源受限 Agent 的规划通常有如下两个重要特征:(1)规划通常是局部规划,而且可以在以后根据实际情况补充这些局部规划;(2)规划通常具有层级结构。例

①Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.49.

如,关于目的的规划,以及关于方法和初始步骤的规划。

由于更一般的意图包含了更具体的意图,因此提前确定这种局部的、层级结构的规划,把更具体的决定留待以后再做的策略,有着深刻的务实理由。规划的局部性和层级结构与规划的惯性相结合,赋予许多意图和行动如下混合特征:一个新的意图或者行动在一个方面可能是慎思的,同时在另一个方面可能不是慎思的。意图或者行动可能是慎思的直接结果,但是这一慎思可能已经把先前意图和规划作为固定的背景,而这些意图和规划在慎思时并不需要重新考虑。

正是通过这种局部的、层级分明的、抵制重新考虑并最终控制行为的规划,使得我们的慎思和行动之间的联系随着时间的推移而不断地延伸。从先前意图到更进一步、更具体的意图,或者关于手段或者初始步骤的进一步意图的推理,规划的局部性都扮演了重要的作用。意图是使得"这些局部规划"和"进行这些局部规划时所需要的推理"得以关联的关键因素。

(二)对规划的要求

规划支持协调,并能够扩大慎思对以后行为的影响。在其他条件相同的情况下,规划需要满足哪些需求才能很好地为这些角色服务呢?首先,规划需要具有内部一致性。为了在一段时间内协调 Agent 的活动,在其他条件相同的情况下,规划应该是内部一致的。其次,规划需要手段与目的具有一致性①。虽然规划通常是局部的,但是随着时间的推移,它们仍然需要适当地补充。认识到规划的这些需求,一方面有助于区分意图和规划,另一方面也有助于区分普通的愿望和评估。

(三)先前规划框架

我们对意图和规划有两个主要的理性要求,与这两个要求相关联的是"意图和规划在实践推理中作为输入所发挥的作用"。首先,考虑到对手段与目的一致性的要求,事先没有重新考虑的意图往往会引起进一步慎思的问题。第二,考虑到意图的一致性,先前意图不需要重新考虑,从而限制了进一步的意图;特别是,它们限制了"手段与目的一致性需求所造成问题的"解决办法。

先前意图和规划为 Agent 决策提供了一个背景框架,并在此框架下进行各种选项的权衡。这个框架有助于 Agent 对如下问题进行慎思:它有助于确定哪些选项是相关的和可接受的。先前意图的作用是帮助确定"在权衡相互冲突的行动理由的过程中"应考虑哪些选项,而不是为"被考虑的替代选项"提供有利的理由。在慎思中需要权衡的理由仍然是愿望-信念理由。通过这种方式,就超越了愿望-信念模型的适度扩展,并把意图直接作为实践推理的输入。

意图为实践推理提供了框架理由,其作用是帮助确定选项的相关性和可接受性;这些框架理由与愿望-信念理由并不重复,而是构建了权衡愿望-信念理由的过程。此外,意图在"为权衡愿望-信念理由提供背景框架方面的作用本身",是基于满足理性愿望的实用性考虑。

Bratman(1987)的意图理论的一个中心问题是"为两种实践推理之间的关系"提供如下令人满意的模型:权衡各种选项的愿望-信念理由,以及从先前意图到关于手段、初始步骤的意图的推理,或者更具体的行动方针②。为此,需要把先前意图视为规划的要素,这些规划提供了一个背景框架,这个框架包括愿望-信念理由的权衡;而且这个框架为进一步的推理提出了问题,并限制了这些问题的解决方案。因此,实践推理有如下两个层次:(1)先前意图和规划会提出问题,并提供对可能解决这些问题的选项的筛选;(2)愿望-信念理由作为考虑因素,在相关选择和可接受选项之间进行权衡。

(四)意图和信念

在对背景框架中意图和规划的作用进行解释时,可以假设存在一种坚定的信念,而不仅仅是置信度在 0 到 1 之间取值的"主观概率"。实践推理和规划通常所依据的背景框架,不仅包括先前意图和规划,而且还包括这些明确的信念。这些心智态度共同构成了推理中解决的决策问题。

现在探讨先前意图和规划是"如何对选项的可接受性进行筛选的"。在进一步的实践推理中,先前意图的这种筛选作用的基础是:Agent 的

①Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, p.61.

②Bratman M E. *Intention*, *Plans*, *Practical Reason*. Cambridge: Harvard University Press, 1987, pp.65-67.

意图与其信念网络需要具有一致性;在其他条件相同的情况下,Agent 应该有可能在一个“他的信念是真实的”世界中做所有他想做的事情。但是,不要假设“这种一致性约束和选项筛选的性质之间存在过于简单的关系”,因为并非所有“与 Agent 已有意图和信念不一致的选项”都是不可接受的。

对于意图和信念的强一致性需求与新选项的可接受性之间关系,可以进行如下理解。考虑一个新的选项 O。固定 Agent 先前意图,并把这些意图添加到 Agent 的意图和“与选项 O 有关的新意图”的信念网络中,而且这个新意图导致了信念的变化,但是 Agent 先前意图并没有导致对信念的任何修改。如果意图和信念网络中的这些变化,不会在该网络中引入新的不一致性,那么选项 O 是可以接受的。

结语

本文从常识性心理学、心灵哲学和行为哲学的视角,系统探讨了作为人类智能、动物智能和机器智能的统一模型的 Agent 的意图、信念、愿望、规划与实践推理的关系,以解释如何从意图的角度来表征 Agent 的心智和行为。主要观点如下:(1)意图既是潜在的行为影响者,又是行为控制者,它为进一步的实践推理和规划提供了重要的输入。(2)意图涉及承诺;规划揭示了“以实践推理为中心的承诺的重要属性”;资源受限 Agent 的规划通常是具有层级结构的局部规划;意图是规划和实践推理得以关联的关键因素。(3)规划支持协调,并能够扩大慎思对之后行为的影响,因此,需要规划具有内部一致性,而且需要其手段与目的具有一致性。(4)意图不仅能够帮助实践推理确定选项的相关性和可接受性,而且还能够构建权衡愿望-信念理由的过程。

至于进一步的研究,可以考虑围绕如下三个方面展开:(1)如何对本文的成果进行形式化?(2)意图、慎思与 Agent 理性的关系如何?(3)意图与 Agent 承诺的关系如何?

The Philosophical Basis for Belief-Desire-Intention Logic

HAO Yijiang

(Institute of Philosophy, Chinese Academy of Social Sciences, Beijing 100732, China)

Abstract: From the perspectives of common sense psychology, philosophy of mind, and behavioral philosophy, a systematic exploration of the relationships between/among intentions, beliefs, desires, plans, and agent practical reasoning can lead to the following main viewpoints: (1) Intentions are not only potential conduct influencers, but also conduct controllers, and they provide crucial inputs for further practical reasoning and planning; (2) Intentions involve commitments. Plans reveal the important properties that are crucial to an understanding of reasoning-centered commitments. The plans of a resource-bounded rational agent are typically partial and have a hierarchical structure. Intentions are key factors in linking planning and practical reasoning. (3) Plans support coordination and systematically extend the influence of deliberation on later conduct, and (4) Intentions not only help to determine the relevance and admissibility of options through practical reasoning, but also structure the process of weighing desire-belief reasons.

Key words: belief; desire; intention; plan; agent practical reasoning

(责任校对　龙四清)