
实现符合约束条件的目标制定和规划

史蒂文-J-琼斯

STEVEN.JONES@CIC.IQMRI.ORG

罗伯特-E-雷

ROBERT.WRAY@CIC.IQMRI.ORG

美国密歇根州安阿伯市 IQM 研究所综合认知中心 邮编：48105

摘要

遵守规范、规则和偏好的一部分是将约束条件（如道德知识）纳入一个人的目标制定和规划处理过程。我们在一个简单的领域中探讨了不同伦理框架中的知识编码如何影响代理的目标制定和规划处理，并展示了当代理的相关约束集合包括各种类型的“硬”和“软”约束时，代理满足和满意的能力。代理如何尝试遵守道德约束取决于道德框架，我们研究了在遵守道德规范时道德框架和功利框架之间的权衡。具有代表性的情景突出显示了在执行同一任务时，对同一规范的不同框架会导致不同的行为。我们的探索表明，元认知判断在解决目标制定和规划过程中的伦理冲突方面发挥着重要作用。

1. 引言

人工代理所处的环境不仅仅由其任务决定。对行为的限制来自很多方面，从物理限制到社会规范。重要的是，在现实世界中，在人类居住的环境中，可能会出现与任务领域并不直接相关（或很容易从任务领域中预测到）的限制因素。相反，领域、任务和整个人类环境（包括法律、规范等）的交汇会引入约束系统，这些约束系统通常包括交互和不一致。例如，代理不可能既满足所有相关约束条件，又按规定执行任务。我们目前的研究重点是识别、研究和应对无数的计算挑战，这些挑战必须得到满足，才能使人工智能代理在符合这些约束条件的同时完成任务（Wray 等人，2023 年）。

人类在制定和评估行动方案时，（通常）会考虑到与自身情况相关的道德规范。或者，也许更重要的是，我们可以说，社会期望他们这样做（而对那些不这样做的人则会给予适当的惩罚，Sripada & Stich, 2006; Molho et al.）此外，即使存在许多相互影响和相互冲突的制约因素，人类也能产生规范可接受的行为（Sripada 和 Stich, 2006 年；Malle 等人，2017 年）。在一个由机器人、虚拟助手、工作队友等各种人工代理组成的世界中，人类将期望这些人工代理遵守规范。- 人类会期望这些人工代理遵守规范，就像他们期望其他人的行为一样。因此，人工代理必须有能力生成符合这些约束的计划和行为（Arkin 等人，2011 年；Rossi & Mattei, 2019 年；Langley, 2019 年；Wray & Laird, 2021 年；Giancola 等人，2020 年；Misselhorn, 2020 年；Malle 等人，2019 年）。

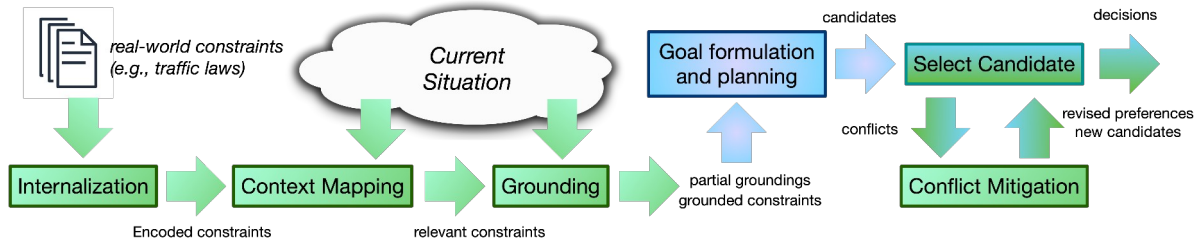


图 1：符合约束条件的计算流程（改编自 Wray 等人，2023 年）。

作为研究和理解人工智能代理如何实现约束条件遵从的更大努力的一部分，本文重点关注代理在计划任务时考虑这些约束条件的能力，以及这些约束条件如何影响代理的任务。我们假定世界具有足够的不确定性（许多制约因素可能来自多个来源）和不稳定性（概念漂移），因此先验训练并不是（完整的）解决方案。相反，在我们的问题概念中，当前的约束条件既影响目标的制定（代理应该完成什么任务），也影响规划（代理应该如何行动来实现目标）。我们强调的是，在存在道德规范的情况下，尤其是那些对行为结果施加“软约束”的规范情况下，代理如何进行规划。我们考虑的这一类规范包括人际礼仪（尊重个人空间）和行为镜像（例如，在图书馆刻意保持安静；开车时与他人保持一致的速度）。

在本文的其余部分，我们将介绍如何设计和实现一个在规划时能够“考虑”规范（和其他约束条件）的代理。我们将通过确定计算要求来更详细地描述这一问题，并引入一个示例领域来比较和对比推理道德规范的其他方法。然后，我们总结了不同设计决策和不同伦理知识对代理行为的影响。重要的是，我们说明了关于如何表示和解释规范的不同假设（“伦理框架”）是如何影响代理的行为和计算需求的。此外，在面临任务与规范冲突的困境时，需要元认知来制定目标和计划。此外，支持这种元认知所需的具体知识内容取决于规划中使用的伦理框架。

2. 背景介绍

Wray 等人（2023 年）提出并描述了一个用于遵守约束条件的计算管道，如图 1 所示。这个端到端的管道是在 Soar（Laird, 2012 年）中实现的。我们为每个核心计算功能（图中方框）开发了“占位符”解决方案。例如，对于内部化，我们开发了一种语言和工具，支持对约束条件的声明性断言，然后将其编码到 Soar 的语义存储器中。该管道旨在进一步探索更复杂的内部计算功能方法。例如，在最近的工作中，我们开发了一种通用算法

表 1: 利用约束系统进行目标制定和规划的要求

| | | 要求简要说明 |
|---------------------------------|----|---|
| 整合 数据流 应软约束和硬约束 | 融合 | 规划必须考虑到不同来源的限制和/或反映不同类型限制的约束。物理限制无法克服（如最高车速），而规则或规范（安全车速、车速限制）则可以克服。 |
| | 适应 | 人类可以处理各种限制，包括严格的禁令、职责和物理障碍（"硬限制"）。人类还可以平衡各种资源限制、"灰色地带"、建议或偏好，而不是具体的禁令。我们可以把这些称为 "软约束"*（Meseguer 等人，2006 年）。目标制定和规划必须考虑 "软" 约束和硬约束的影响。 |
| 预测互动 和失败。 | 和 | 代理不能假定与其处境相关的约束条件是一致或连贯的，也不能假定它们不会与其任务相冲突。在许多情况下，规划无法满足所有约束条件和任务。规划算法必须具备预测和应对失败的机制。 |
| | | 代理的算法必须对环境变化做出反应。 |
| 反应灵敏，随时 随地 适应部分可 观察性 识别学习 | | 代理不能假定其感知足以观察到与其规划相关的所有变量或状态。 |
| 在线、渐进式学 习 | | 代理的规划性能应随着领域经验的积累而提高。一般来说，随着对问题解决 的熟悉，处理过程应从深思熟虑 ("系统 2") 转变为识别 ("系统 1") (Kahneman, 2011 年)。由于代理必须对环境做出反应（也可能对环境本身 的动态变化做出反应；漂移），因此代理应在执行过程中学习 ("在线")。 |

* 作为一个技术术语，"软约束" 可以指一种特定的形式主义，但我们只是通俗地使用它。

这对于评估潜在未满足的约束条件（例如，当约束条件只能部分基于当前情况时可能出现的情况）来说是必要的。本文的重要观点是，这种 "约束遵从" 管道为探索端到端代理行为（包括推理约束和执行任务域中的任务）提供了一种工具，同时使研究重点能够放在管道内的个人能力上。

如上一节所述，我们所报告的研究主要集中在管道的目标制定和规划部分。这个方框在图中使用了不同的颜色，因为一般解决方案需要同时考虑一般约束条件和当前任务的具体要求。因此，我们的研究目标是确定（算法上的）约束条件应如何影响目标制定和规划。显然，规划算法在规划时已经考虑到了约束条件。我们认为，实现将道德约束考虑在内的规划这一挑战凸显了三个关键要求。表 1 的前三行总结了这些要求。虽然我们在本文的其余部分强调了这三项要求，但其他要求也会影响后面的设计和实现。表 1 的其余部分

列出了其他要求，并作了简要概述。其中许多要求源于 "约束合规性 "的一般要求，Wray 等人（2023 年）对此进行了进一步讨论。

3. 实验领域和任务

为了更具体地说明这些要求，以及在制定目标和规划时面临的挑战，本节将介绍一个领域、任务，以及该领域可能出现的潜在制约因素的示例。

示例域是一个海运仓库。仓库储存不同类型的货物。在实际的海运仓库中，存在物理限制（如各种货物的尺寸、重量、属性和位置）、法律和正式程序的限制（如《美国法典》第 2024 条）以及影响人际互动的规范。例如，在一个既有人又有机器人的仓库中，我们假定机器人一般应遵守与在他人周围进行身体移动相关的规范。在一个大型仓库中，除非你打算以某种方式与其他人接触（检查他们的身份信息、询问货物的状态等），否则你通常不会与其他人靠得很近。

我们设想的机器人代理可以在这一领域执行许多不同的任务。例如，机器人可以执行各种安全功能（如巡逻、检查人员身份、护送人员等）。在本文中，我们只关注一项任务：巡逻。机器人应定期环绕房间四周巡逻。在某些海运仓库环境中，这种巡逻是例行工作，也是必需的（《美国法典》，2024 年）。

图 3 是我们对这一领域的模拟描述。与其他研究人员（Loreggia 等人，2022 年）类似，我们采用了相对简单的网格来表示仓库环境。虽然这只是对实际仓库复杂性的粗略简化，但环境中既有物体也有人，可以支持许多任务和多种类型的约束条件。

在巡逻任务中，理想的情况是机器人以最快的移动速度访问周边的每一个单元。不过，我们在定义任务时有两个最低性能标准：1）机器人应访问周界单元总数的某个特定部分；2）机器人必须在不超过规定时间内完成巡逻。（为了进一步简化规划，我们用移动次数来代表时间：移动到一个开放的小区需要消耗一个单位的时间）。

环境中的物体会造成不同类型的限制。箱子（棕色立方体）是一种物理限制：机器人无法进入这些单元。桶（蓝色/橙色图标）是一种物理限制，允许移动，但速度较慢。在本文中，我们假定穿越装有木桶的牢房所需的时间是穿越开放牢房所需的时间的两倍。

除了物理限制外，我们还引入了以下规范：代理应避免进入有人居住的牢房。这一准则可能代表着礼貌或对他人的尊重。考虑到这一准则，规划的结果通常是选择穿过有人的牢房，以提高巡逻效率，或者绕过有人的牢房。本文的其余部分将讨论如何在规划算法中体现这一约束条件。

在接下来的例子中，我们将对两类情景下的行为进行对比，我们称之为 "共识 "和 "两难"。共识情景旨在呈现一个明显的

或容易的情况，在这种情况下，无论机器人如何解释规范，都不会对其应走的适当路径产生歧义。图 3(a) 就是一个例子。相比之下，两难情景（如图 3(b)）的设计目的是让其他解决方案显而易见。机器人选择哪种方案取决于机器人（或机器人设计者）的假设。

4. 根据道德约束进行规划

在本节中，我们将介绍代理如何在规划任务时考虑各种约束条件。我们将看到，规范需要一定的框架来解释和评估，而框架的选择会从本质上改变代理在相同规范下的行为方式。我们将介绍一种简单的规划方法，然后讨论如何对其进行阐述，以实现不同类型的伦理框架，从而纳入规范。

4.1 基准规划算法：A*

我们使用 A* 搜索来实现导航规划。我们选择 A* 有三个原因：1) 它为人所熟悉，研究（和演示）的重点是如何纳入规范（而不是为这项任务开发复杂的规划解决方案）；2) 我们计划对比不同的方法，这意味着不需要优化绝对性能；3) Soar 在其 "默认知识" 中提供了现有的 A* 算法，从而加快了实施速度（我们或多或少可以立即使用 A*）。

算法 1 概述了 A* 的实现。方框内的文字代表我们为适应规范而修改的 A* 部分（将在下一节讨论）。一般来说，A* 会迭代评估搜索节点，计算每个节点的估计总成本 f 。 f 是该节点的成本 g 与从该节点到目标的可接受启发式估计成本 h 之和，其中 $f = g + h$ 。

为了让 A* 适应巡逻任务（起点和终点位置相同），我们开发了一种自定义启发式，而不是使用通常用于网格世界的曼哈顿距离。对于巡逻任务，搜索中的每个节点代表的不仅仅是一个具有给定位置的单元格。相反，每个单元既代表了迄今为止周边单元的覆盖范围，也代表了该单元的位置。向目标状态 "靠近" 意味着要覆盖更多的边界单元。我们将当前节点状态表示为一个向量： $s_c = [x_c, y_c, n_c]$ ，其中 n_c 是截至目前已覆盖的边界单元格数。¹ 目标状态是一个向量： $s_g = [x_g, y_g, n_g]$ 。可接受的启发式函数是分段定义的，以处理边缘情况，但核心概念是低估覆盖剩余所需边界单元并返回起点的预期移动次数： $h(s_c, s_g) =$

$\max(n_g - n_c, \text{ManhattanDistance}([x_c, y_c], [x_g, y_g]))$ 。

4.2 调整规划以适应规范：替代性伦理框架

在我们的探索中，我们重点关注对相关工作中作为重点的两大规范性伦理学类别的支持：道义论和功利主义。功利主义一般是指选择任何能使整体 "幸福" 或效用最大化的行动 (Mill, 1998 年)。道义论

1. 此外，我们还可以跟踪已穿越的特定边界细胞，但在此我们将简化表述。

算法 1 A* 的伪代码，改编自维基百科 (2024)。

```

1: aStar (起点, 终点)
2: openSet = priorityQueueSet({}); cameFrom = map({}); g = map({}); f = map({});
3: g[start] = 0
4: f[start] = h(start)
5: openSet.add(start, f[start])
6: 当 openSet 不为空时 do
7:   current = openSet.top()
8:   如果 当前是目标, 那么
9:     return reconstructPath(cameFrom, current)
10:  openSet.pop()
11:  for adjacentCell in current.options() do
12:    如果 向相邻单元格的移动是 法律 则 # 道德观改变了 "合法性"
13:      tentativeG = g[current] + cost(current, adjacentCell) # Util.
14:      如果 (adjacentCell 不在 g 中) 或 (tentativeG < g[adjacentCell]) 则
15:        cameFrom[adjacentCell] = current
16:        g[adjacentCell] = tentativeG
17:        f[相邻单元格] = tentativeG + h[adjacentCell] (暂定 G
18:        如果 adjacentCell 不在 openSet 中, 那么
19:          openSet.push(adjacentCell, f[adjacentCell])
20: 返回 失败

```

这通常是义务论的特点，即禁止选择被认为是 "坏" 的行为，即使这可能不会导致功利主义的结果 (Korsgaard, 1986 年)。我们从根本上遵循了对义务论的这一解释。然后，我们不再把我们的规范固有地归结为一种禁止，或者是代表侵犯他人空间的一种代价，而是把这两种可能性都考虑在内。

4.2.1 功利主义

当机器人与人类共处一室时，我们可以把避开他人的准则理解为需要付出代价。这种选择在功利主义框架中很常见。例如，Loreggia 等人 (2022 年) 通过在路径规划中使用的移动成本之外添加成本，实现了类似的规范。功利主义方法通常使用成本来表示各种约束条件，这样就可以相对直接地将各种约束条件整合到一个共同的决策框架中。

为了将这一准则纳入功利成本，我们改变了算法 1 第 13 行中的成本函数。当前的成本函数在总成本中加入了从当前位置到下一个目的地的移动成本。如果目的地有人，则成本计算为移动成本与用户定义参数之和，该参数代表进入他人私人空间的成本。一般来说，Soar 支持对其工作存储器中可用于计算此类成本的状态特征进行任何评估。

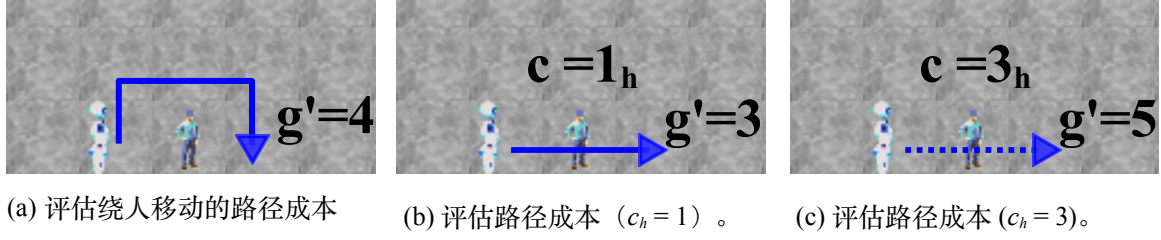


图 2：计划路径取决于移动成本和带人进入小区的成本。

图 2 是这种效果的一个简单表示。在左图 (a) 中，代理人计划绕着一个人移动时的成本为 4。在中间 (b) 处，当代理计划穿过一个人时（成本_{human} = 1），穿过一个人的总成本为 3，因此穿过一个人比绕过一个人更便宜。²在 (c) 中， $c_h = 3$ ，通过被占单元移动到指定位置的总成本为 5。在规划整个地图的巡逻路径时，累积效应可能会更加复杂，因为在边界单元覆盖和多人和/或多物体的情况下，路径选择比评估单个偏差更加复杂。

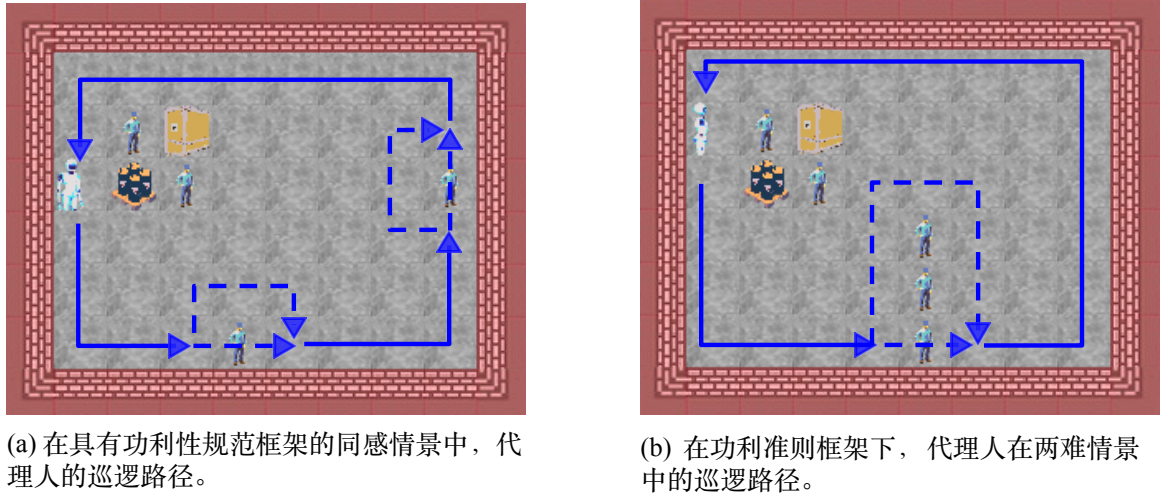


图 3：在共识情景和两难情景中，当使用个人空间规范的功利框架时，代理人的巡逻路径用蓝色箭头表示。虚线箭头表示代理人计划和执行的路线取决于与侵犯个人空间相关的编码成本。

在共识情景中，正确的巡逻路线在设计上是直截了当的。图 3a 显示了在将道德规范解释为反映功利成本的情况下，代理人执行巡逻路线的共识情景。

2. 请注意，在本例及整个例子中，我们假定进入一个单元格的成本是根据该单元格的属性累加的；因此，包含该人的单元格的成本为 $C_{total} = C_{empty} + c_h$ 。

在进退两难的情况下,行为主体需要在大幅延长巡逻路线以避免人群,或选择一条经过人群的较短路线之间做出选择。图 3b 显示了代理人在两难情况下的路径,其中成本 > 8 会使代理人避开人群。

这种行为是预料之中的。在某些情况下，将避免干扰他人视为一种礼节是合理的。但是，当这样做的成本过高时，尤其是在完成另一项任务的情况下，就不应该强制执行这种礼节。在使用 A* 计划时，将适度的成本与通过人群联系起来，自然会产生这种行为。

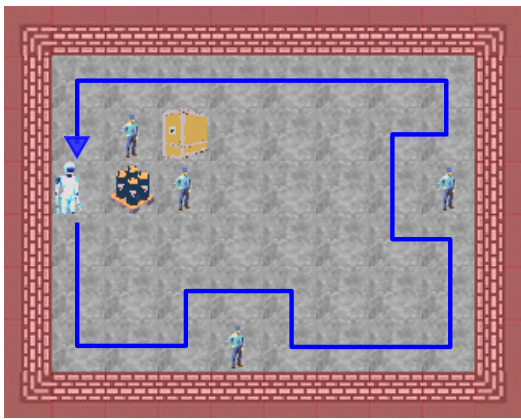
4.2.2 道义论

另外，我们也可以把“避开他人”这一准则理解为一种硬性的职责约束，即代理人必须避免进入有人的牢房。在我们的代理中天真地执行这一点，就意味着严格拒绝将进入有人的牢房作为计划期间的考虑选项。

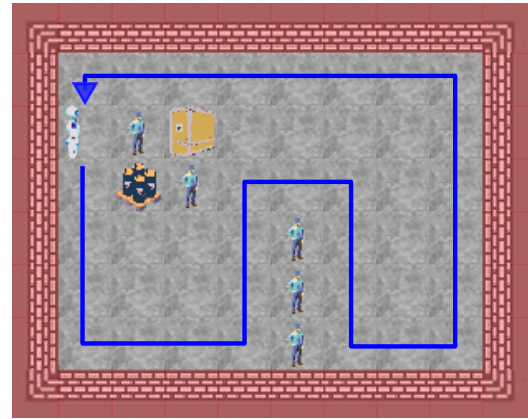
为了纳入被解释为道义禁止的规范，我们改变了代理作为 A* 边界的一部分所考虑的选项。算法 1 中的伪代码在第 12 行检查了扩展 A* 搜索的节点是否合法（例如，它不能穿过墙壁）。（例如，它不能穿过一堵墙。）对于道义论，我们在第 12 行增加了这一检查，以检查单元格中是否包含一个人。如果是，则跳过。一般来说，Soar 支持任何对状态特征的评估，这些状态特征可以在其工作记忆中表示为这样的禁止。

图 4a 显示的是在将道德规范解释为去道德禁止的情况下, 代理人执行巡逻路线的共识情景。图 4b 显示了代理人巡逻路线的两难情景。请注意, 它从未计划或执行过穿过人群的选项, 即使计划的路线成本很高。

这些行为是意料之中的。不过，请注意，我们在设计场景时已经考虑到了代理可以取得成功。在巡逻任务中，如果没有为额外的移动留出很大的余地，那么采取



(a) 代理人在同意方案中采取的巡逻路径。



(b) 在两难情景中，代理人的巡视路径是以 "义务" 为框架的。

图 4: 在共识情景和两难情景中, 当使用个人空间规范的去道德框架时, 代理人的巡逻路径用蓝色箭头表示。

如果绕着人走很长一段路，巡逻就不可能成功。在两难的情况下，使用 "去义务论" 框架意味着代理人的巡逻总是会失败。

5. 通过元认知解决冲突

在共识情景中，无论代理人使用的是具有足够成本的去道德框架还是功利框架，规划的路线都是相同的。在两难情景中，去道德框架意味着代理人总是计划绕过人群，但功利框架只有在干扰人群的成本比共识情景中更高时才会绕过人群。

然而，我们最初的实施方案在某些情况下可能会出现不良行为。在功利框架下，当 c_h 4 时，即使代理已将规范编码为知识，并检测到它违反了该规范，代理仍会计划与人一起进入边界牢房。这

即使避免违反规范（相对于巡逻的总成本而言）是微不足道的，"违反规范" 也会发生。虽然这是成本形式主义的逻辑结果，但它可能不是可取的行为。同样，如果侵犯个人空间是完成巡逻的唯一选择，那么在可能接受侵犯个人空间的情况下，违背道义的代理人可能不会完成巡逻。作为回应，我们扩大了实施范围，将规划期间的元认知包括在内，以支持更灵活地将伦理知识应用到实际情况中。

5.1 与道德规范冲突的知识

根据我们对任务的定义，完成巡逻的一部分就是少于允许的最大移动次数。以图 4b 中的代理路径为例。这个代理用 30 次移动完成了一次巡逻。然而，假设一个代理必须在不超过 29 次移动的情况下完成一次巡逻。这将不是一个可行的计划。在这种情况下，一个以避免干扰他人为准则的代理人就会面临选择。它要么完成巡逻并违反道德规范，要么尊重道德规范并完成巡逻任务。

我们认为这是一个例子，说明在有额外知识的情况下，可以利用这些知识作为对计划失败的元认知反应。假设成功巡逻至关重要，而避免干扰则不那么重要。在这种情况下，我们认为避免干扰他人的道德准则不应消失，而应放松到足以完成巡逻的程度。这样，代理人就会努力履行其职责，但可以优先考虑更重要的职责。

在修改后的实施方案中，当代理在避开人的同时找不到合适的计划时，它可以承认多干扰了一个人。因此，当代理之前因为图 4b 中的路线需要过多移动而陷入困境时，代理可以规划最多干扰一个人一次的路线：最终根据图 3b 中经过一个人的路线进行规划和执行。

5.2 与功利准则冲突的知识

另一种可能是，功利代理人似乎成功地规划了一次巡逻，但却计划扰乱他人。回想一下，在扰乱人们的成本很低的共识情景中，功利主义者

即使相对于整个巡逻的成本来说，避开这些人的成本并不高，但特工还是会扰乱人们的视线（即使只有一个人挡路，特工也会扰乱别人的视线）。(即使只有一个人挡路，它也会扰乱别人)。

对规范的了解会对行为产生影响。如果一个代理能够发现自己知道某个规范，但与该规范相关的成本却微不足道，那么就有理由重新评估与干扰他人相关的编码成本。

在修改后的实施方案中，当代理规划的路线进入被占用的单元格时，会进行一些额外的计算。首先，它会测试 $c_h \geq 4$ 。如果是，它就会评估自己是否有额外的行动空间

(即可以完成巡逻任务)。如果有，则代理重新扫描 $c_h = 5$ 。

(成本大于 4 有可能导致绕过边境上的人的路径比干扰人的路径成本更低，这反映了规范有可能影响巡逻的执行)。如果得出的计划仍能满足巡逻要求，则选择该计划。在共识情景中，元认知的使用会导致代理人在 $c_h = 1$ 时选择如图 4a 所示的路径（即图 3a 所示的路径）。此外，两难情景下的计划路线没有变化。

6. 结果与讨论

我们现在介绍总体结果，包括相对计算成本的比较，以及该方法的能力和局限性的进一步描述。

表 2 总结了当前实施方案的行为和性能。请注意，不同的代理配置适用于不同的场景。共识情景中省略了 "道义与元认知" 配置，两难情景中省略了 "功利与元认知" 配置，因为在每种情况下，元认知都不会改变行为。共识情景中的功利框架 ($c_h = 2$) 显示了 c_h 设置过低对巡逻行为的影响。在两难情景中， $c_h = 5$ 的提出是因为在共识情景中， $c_h > 4$ 总是会导致特工绕开人群；而在两难情景中，选择略高于此临界值的 c_h 有可能会影响特工的行为。同样，之所以选择功利主义 ($c_h = 9$)，是因为任何 $c_h > 8$ 都会导致代理人绕开他人。

表 2：代表性场景/代理配置的代理性能汇总。

| 共识方案 | 道义论 | 功利主义 ($c_h = 2$) | 功利主义 ($c_h = 9$) | 功利主义 ($c_h = 2$) + 元认知 |
|--------|------|--------------------|--------------------|--------------------------|
| # 决定 | 5055 | 587 | 5915 | 6928 |
| 避开人群 | 真 | 错误 | 真 | 真 |
| 巡逻取得成功 | 真 | 真 | 真 | 真 |
| 两难困境场景 | 道义论 | 功利主义 ($c_h = 5$) | 功利主义 ($c_h = 9$) | 道义论 + 元认知 |
| # 决定 | 3639 | 1448 | 14654 | 3913 |
| 避开人群 | 真* | 错误 | 错误 | 错误 |
| 巡逻取得成功 | 错误 | 真 | 真 | 真 |

* 特工避开了人们，因为它一动不动。

在进退两难的情况下（假定有足够的行动分配）。这也表明，即使行为没有改变，成本参数也会对性能产生影响。

我们总结了几个定性概括：1) 功利主义框架使代理人有时会认为穿过人群的路线成本较低，即使有可能避开人群并满足巡逻要求，也会选择这些路线；2) 当可以避开人群时，义务论的计算成本较低，但当无法避开人群时，义务论的计算成本较高，3)

$c_h \gg c_{threshold}$ 的功利主义会导致与去道义禁止相似的行为（除了最终可以找到一个通过人的解），4) c 的具体值是多少？_h

5) 元认知可以为代理处理编码规范提供更大的灵活性，并有可能使代理的行为更好地符合人类的期望。

关于结果 2 和 3：在达成共识的情况下，道德框架和功利框架（具有足够高的成本）的主要区别在于，代理人使用道德知识规划移动所需的认知周期更少，因为需要考虑的可能路线更少。对规范有道德认知的行为主体平均需要花费 5055 个认知周期，而对规范有功利认知的行为主体平均需要花费 5915 个认知周期。我们预计这种处理过程中的质的差异是普遍存在的，因为即使功利主义导致避开人群，一些进入有人居住的小区的路线仍会被评估。

除了已经提供的演示之外，我们还探索了可纳入代理的潜在规范空间，以及道德约束和其他约束之间的潜在相互作用。如第 2 节所述，一些测试场景包括桶（2 倍于开放单元格的移动成本）和板条箱（阻碍）的约束条件。图 5 显示了包含这些额外限制的计划移动。

从设计上讲，这些额外的约束相当于硬性禁止和额外成本，类似于道义上的禁止和效用。然而，它们并不能提供所有相同的考虑因素。

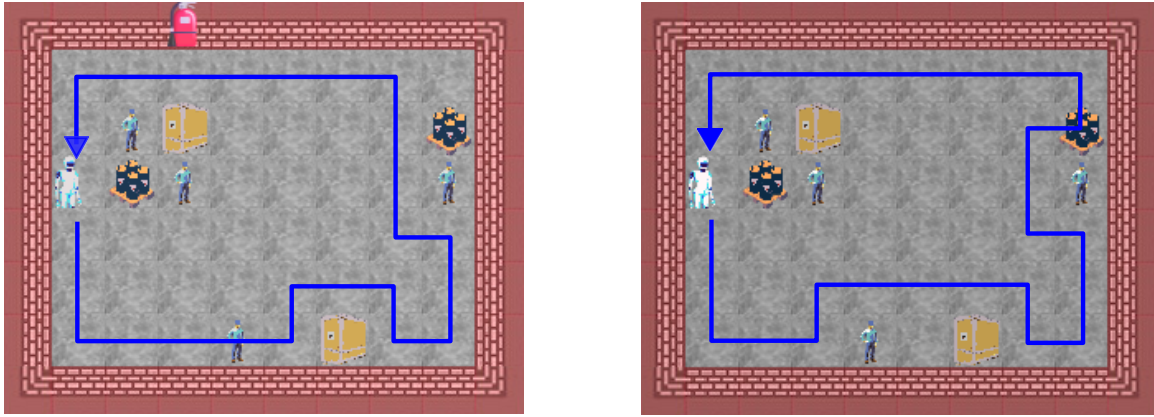


图 5：这个代理使用了道德约束的功利框架，并包含了对板条箱和木桶约束的敏感性。当个人空间成本为 2 时，代理规划的路线在左边；当个人空间成本为 6 时，代理规划的路线在右边。

代理不能决定忽略通过板条箱的约束。代理人不能决定无视关于通过箱子的限制，木桶的成本应反映对移动难度的实际估计，而不是对遵守社会规范的执行程度。

因此，迄今为止的研究可以帮助我们进一步探究在有约束条件的情况下制定计划时产生的其他元认知决策。与规范/道德约束相比，物理约束的成本通常更容易观察/感知。如果成本函数是对时间或精力的度量，那么如果“开放地板”需要一个单位的时间/精力，而“木桶牢房”的成本是“开放地板”成本的2倍，那么这个成本就可以直接/有效地通过环境中的互动来确定。然而，估算违反规范的（适当）成本并不简单（Malle 等人，2019年）。估算规范成本的“噪音”可能会促使人们采用去道德主义方法（即在缺乏准确成本的情况下）。一般来说，在某些情况下，行为主体本身可能需要确定最合适的伦理框架/策略，或许需要思考框架对其自身行为的影响。在这种情况下，行为主体在不知道违反准则的代价时，可能会选择默认的义务论框架。

A* 规划的一个缺点是，我们在学习过程中并没有展示出显著的迁移性或通用性。代理可直接对网格单元表征进行推理，并通过 Soar 分块，重新创建可替代规划的识别知识。然而，在当前的实现过程中，代理并没有将学到的知识迁移到类似的情境中，因为代理并没有对抽象概念进行推理，而这些抽象概念可以在不同的情境中通用。一个潜在的设计决策（放弃 A* 的最优性保证）是使用能对抽象表征进行推理的规划。Soar 支持用额外的抽象表征来增强工作记忆，并将其作为推理的基础，但这一点在本研究中还未涉及。

当前系统的另一个特点是，它既能在计划过程中表示对失败的预期，又能在执行过程中检测失败。对失败的预期为元认知评估提供了基础，即有可能放宽约束条件。考虑一下代理无法单独成功的更困难的情况，比如一排箱子完全堵住了半个房间。在这种情况下，具有领域知识的人类很清楚，仅靠代理是无法完成巡逻任务的（例如，可能需要一辆叉车来移动部分板条箱）。在目前的实现中，代理最终会认识到在这种情况下它无法成功。然而，提高代理效率的一个潜在方法是将任务或特定领域的知识融入其中，以便在代理的能力范围内更快地识别出“不可能”的情况，而不需要穷尽搜索。虽然这种能力可以融入到代理中，但我们还没有探索过针对特定任务的知识工程改进，也没有学习过检测这种情况的特征。

回顾表 1，当前的代理可以满足大多数要求。可以同时处理多种冲突类型，包括来自不同来源、意味着不同承受能力的限制。Soar 的潜在元认知能力与在线处理冲突的能力非常吻合。虽然这不是本文的重点，但我们已将 A* 调整为随时 A*，以提供计算受限的响应。此外，现有的代理在部分可观察性方面的表现也是可以接受的，基本上是为它能观察到的房间部分规划出一条最佳路径，然后移动，再停下来重新规划。

7. 相关工作

这项工作有助于实现在自主系统中整合规范约束的更大目标。我们工作中的一个明确假设和方法限制是，在制定目标和规划时要考虑到所有符合约束条件的要求（即表 1）。在本节中，我们将简要回顾相关工作，并与我们的方法进行比较/对比。

Loreggia 等人（2022 年）探索了在网格世界中导航时的规范决策。他们的目标是了解决策空间，并评估性能环境外的导航选择（将这些选择与人类决策进行比较）。我们的代理在性能环境中进行规划和行动。我们尚未将我们的代理计划与人类在相同情况下的计划/路线进行比较。不过，虽然我们没有将其作为表 1 中的一项要求，但代理应该做出人类认为“可接受”的决策。因此，未来与他们类似的比较适合本研究。

Svegliato 等人（2021 年）探讨了将伦理理论与任务决策相结合的问题。他们描述了不同的实施方法，其中每种实施方法都支持一种伦理框架。这种方法可以进行类似表 2 所示的比较。我们概述的综合方法的一个优势是，代理可以同时考虑多种形式的约束知识，并（长期）在不同的伦理框架之间切换。不过，我们只支持一种规划算法的简单扩展，在该代理中全面（或至少更全面）开发各种伦理框架仍是未来的工作。Awad 等人（2020 年）考虑了如何通过计算支持同时反映道义论和功利主义的伦理知情偏好。他们强调，代理的偏好认知表征需要包含上下文信息和行动评估基础，这与我们在目标中对 A* 的元认知扩展类似。有趣的是，他们将“去义务推理”（deontological reasoning）和“功利推理”（utilitarian reasoning）分别映射到系统 1 处理（“硬性”规则）和系统 2 处理中。与此相反，我们的实施方案将“道义论”和“功利主义”视为不同形式的伦理知识，而不是不同处理过程的固有参照（Kahane, 2012）。这种方法可以在新的/不熟悉的情况下进行深思熟虑的推理和规划，并在以下情况下采取更多识别行动熟悉的情况。

我们的工作与伦理治理者（Arkin et al.）我们同样考虑了义务论和结果论伦理学以及对设计选择的计算探索。与伦理治理者不同的是，我们设想的约束遵从是一种终身认知的普遍能力，而不是在任务间重置的模块。要求包括在线增量学习和识别学习，而其行为适应（“内疚”）似乎并不能通过改变代理知识来促进持续适应，也不能提高效率。

Lu 等人（2024 年）描述了如何调整 POMDP，使其包括对道德约束的敏感性。他们的系统同时执行校园巡逻任务（与我们的任务类似），并遵守亚里士多德的美德伦理。他们不支持多种伦理框架，也没有解决伦理知识的在线学习或适应问题。我们的方法未来的一个扩展可能是将其扩展到其他伦理框架，如美德伦理。

8. 结论

考虑到之前介绍的目标和要求，迄今为止的主要局限是我们只探索了一种目标制定和规划算法（A*）。不过，通过这种探索，我们对道德约束如何影响目标制定和规划的内容和（计算）成本有了更深入的了解。此外，我们还期望这一原型的许多方面都能得到推广。对于以效用为导向的规范，我们希望以从观察到的环境特征中得出的预期成本的形式纳入伦理知识。对于禁止性的道义责任，我们希望代理在规划过程中拒绝某些选择，除非其他知识认为有必要这样做。对于纳入道义上的义务，我们希望将伦理知识作为额外的目标（例如，在以前的工作中，以安全为导向的约束是有基础的，Jones & Wray, 2023）。当计划已经存在但似乎不尽人意或未能创建时，元认知推理可以进行干预，以修改目标、计划期间的可用选项或编码规范的方式（例如，改变成本）。在本次调查中，A* 为生成结果提供了高效的实现方法。尽管如此，今后的工作仍将尝试确定更多原则，以便将伦理框架普遍纳入规划，尤其是针对更真实的开放式领域。

未来的工作还将研究更逼真的具体场景的效果，例如有时间压力和部分可观察性的场景。我们认为，类似 Lu 等人（2024 年）的探索（但还包括有限时间和与其他在线认知处理的整合）对于实现符合约束条件的目标制定和规划处理更加适用于现实世界环境非常重要。在线代理通常必须克服在知识不全和时间有限的情况下确定如何遵守各种约束的困难。

本文为在人工系统中实现类似人类的伦理决策这一更大挑战做出了贡献（Rossi & Mattei, 2019）。我们的重点是在传统规划中整合伦理知识。除了前面讨论的具体成果和见解外，由此产生的实施方案还为进一步了解如何将更多种类的伦理知识整合到在线决策代理的认知中提供了一个试验平台。这可能包括更复杂的功利主义和/或道义论概念，比如包含决策理论的道义论（Lazar, 2017）。此外，未来的工作还可以探索纳入更多伦理框架。基于典范的美德伦理（Zagzebski, 2010 年）有可能与基于案例的推理很好地结合起来，后者可以从记忆中检索典范代理人在某一领域的行为方式的相关知识（Kuipers, 2016 年）。其他形式的美德伦理（Annas, 2016 年，第 2 章）可以通过元认知来实现，元认知会评估一个人的计划在多大程度上使其能够保持特定的美德，这些美德是一个人自传体记忆中自我概念化的特征。许多规范可能需要与心智理论（目前尚未实施）相结合的元认知。然而，由于在该方法中，约束遵守的实施与认知更普遍地结合在一起，我们假设这些道德理论中使用的其他必要知识可以作为程序记忆、语义记忆和情节记忆的额外内容来实现，因此我们认为在这一贡献的基础上，这一方向是一个很有前景的方向。

致谢

这项工作得到了海军研究办公室（Office of Naval Research）N00014-22-1-2358 号合同的支持。本文件中的观点和结论仅代表作者本人，不应被解释为代表国防部或海军研究办公室明示或暗示的官方政策。美国政府有权为政府目的复制和分发重印本，尽管本文有任何版权说明。感谢 Peter Lindes 对仓库模拟的贡献。

参考资料

- Annas, J. (Ed.).(2016).《发展美德：综合视角》。New York NY：牛津大学出版社。
- Arkin, R. C., Ulam, P., & Wagner, A. R. (2011). 自主系统中的道德决策：执行、道德情感、尊严、信任和欺骗。 *Proc. of the IEEE*, 100, 571-589.
- Awad, E., Levine, S., Loreggia, A., Mattei, N., Rahwan, I., Rossi, F., Talamadupula, K., Tenenbaum, J., & Kleiman-Weiner, M. (2020). 何时打破规则在道德上是可接受的？基于偏好的方法。第12届 *Preference Handling* 进展多学科研讨会。IOS Press.
- Giancola, M., Bringsjord, S., Govindarajulu, N. S., & Varela, C. (2020). 不确定性条件下自主代理的伦理推理。 *Intrnl Conf. on Robot Ethics and Standards (ICRES)* (pp. 1-16).台北。
- Jones, S., & Wray, R. E. (2023). 约束遵从中部分接地的挑战》。 *Proc. of the 2023 Conf. on Computation Science and Computational Intelligence (CSCI)*.Las Vegas: CPS.
- Kahane, G. (2012).On the wrong track：道德心理学的过程与内容。 *Mind & Language*, 27, 519-545.地点：英国 Wiley-Blackwell Publishing Ltd. Publisher: Wiley-Blackwell 出版有限公司。
- Kahneman, D. (2011).《思考，快与慢》。纽约：纽约：Doubleday.
- Korsgaard, C. M. (1986).说谎的权利：Kant on dealing with evil. *哲学与公共事务*，（第325-349页）。
- Kuipers, B. (2016). 机器人为何以及如何才能符合道德规范？ *社交机器人能做什么、应该做什么：机器人哲学 2016/TRANSOR 2016 会议录*，《人工智能与应用前沿》第290卷。IOS Press.
- Laird, J. E. (2012).《翱翔认知架构》。马萨诸塞州剑桥市：麻省理工学院出版社。
- Langley, P. (2019).可解释的、规范的和正当的代理。 *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 9775-9779).夏威夷：AAAI Press.Number: 01.
- Lazar, S. (2017).Deontological Decision Theory and Agent-Centered Options. *伦理学*，127, 579- 609。
- Loreggia, A., Mattei, N., Rahgooy, T., Rossi, F., Srivastava, B., & Venable, K. B. (2022).Making Human-Like Moral Decisions.2022 年 *AAAI/ACM 人工智能、伦理学会议论文集*，

- 与社会》(第 447-454 页)。Oxford United Kingdom: ACM.
- Lu, Q., Svegliato, J., Nashed, S., Zilberstein, S., & Stuart, R. (2024). 部分可观察性条件下符合道德规范的机器人系统。 *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*. 日本横滨。
- Malle, B. F., Bello, P., & Scheutz, M. (2019). 具有规范能力的人工代理的要求。 *2019 AAAI/ACM 人工智能、伦理与社会会议论文集》* (第 21-27 页)。New York, NY, USA: Association for Computing Machinery.
- Malle, B. F., Scheutz, M., & Austerweil, J. L. (2017). 人类和机器人代理的社会和道德规范网络》。 *有机机器人的世界：机器人伦理国际会议：ICRE 2015* (pp. 3-17). Cham: Springer International Publishing.
- Meseguer, P., Rossi, F., & Schiex, T. (2006). 软约束。 In F. Rossi, P. Van Beek, & T. Walsh (Eds.), *Handbook of constraint programming*, Foundations of artificial intelligence. 阿姆斯特丹
波士顿: Elsevier, 第 1 版。OCLC: ocm70408044。
- Mill, J. S. (1998). *Utilitarianism*. Oxford Philosophical Texts. 牛津大学出版社: 牛津大学出版社。(原著出版于 1861 年)。
- Misselhorn, C. (2020). 具有道德能力的人工系统? 研究设计及其在老年护理系统中的实施。 *人工智能》*, 278, 103179。
- Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). 日常生活中对违反规范行为的直接和间接惩罚。 *自然通讯》*, 11, 3432。
- Rossi, F., & Mattei, N. (2019). 构建有道德约束的人工智能。 *Proc. of the Thirty-Third AAAI Conference on Artificial Intelligence* (pp. 9785-9789). 美国夏威夷檀香山: AAAI 出版社。
- Sripada, C., & Stich, S. (2006). 规范心理学框架》。 In P. Carruthers, S.S. Laurence, & S. Stich (Eds.), *Innateness and the structure of the mind*, Vol. 牛津大学出版社: 牛津大学出版社。
- Svegliato, J., Nashed, S. B., & Zilberstein, S. (2021). 符合伦理的顺序决策。 *美国人工智能学会会议论文集》*, 35, 11657-11665。
- 美国法典》(2024 年)。航行和通航水域。
- 维基百科撰稿人 (2024 年)。 A* 搜索算法 -- 维基百科, 自由的百科全书。 https://en.wikipedia.org/w/index.php?title=A*_search_algorithm&oldid=1214713509. [Online; accessed 8-April-2024].
- Wray, R. E., Jones, S., & Laird, J. E. (2023). 通用智能约束条件的计算级分析》(Computational-level Analysis of Constraint Compliance for General Intelligence)。 *Proc. of Artificial General Intelligence (AGI) Conf.* Stockholm. 斯德哥尔摩。
- Wray, R. E., & Laird, J. E. (2021). 将抽象行为约束纳入代理任务执行。 *Proc. of International Conf. Intelligence*. Las Vegas, NV: Springer.
- Zagzebski, L. (2010). Exemplarist Virtue Theory. *Metaphilosophy*, 41, 41-57.