arXiv:2303.01711v2 [cs.AI] 5 Aug 2023

NovPhy：开放世界环境中的物理推理测试平台

Chathura Gamage[a, c], Vimukthini Pinto[a, c], Cheng Xue[a, c], Peng Zhang[a], EkaterinaNikonova[a], Matthew Stephenson[b], Jochen Renz[a]

[a] *School of Computing, The Australian National University, Canberra, Australia*
[b] *College of Science and Engineering, Flinders University, Adelaide, Australia*
[c] *These authors contributed equally*

摘要

随着能够与物理环境交互的人工智能系统不断涌现，将物理推理能力融入这些系统引起了越来越多的关注。但要在真实的物理环境中运行，仅具备物理推理能力就够了吗？在现实世界中，我们总会遇到从未经历过的新颖情境。作为人类，我们能够有效地适应这些情境。同样，智能体也需要具备在新颖性影响下运行的能力，才能在开放世界物理环境中正常运作。为促进此类人工智能系统的发展，我们提出了一个新的测试平台 NovPhy，要求智能体在存在新颖性的情况下对物理场景进行推理，并据此采取动作。该测试平台包含一些任务，要求智能体检测物理场景中的新颖性并适应它们。为了创建测试平台中的任务，我们设计了八种新颖性，以构成多样化的新颖性空间，并将其应用于物理环境中五种常见的场景，这些场景涉及施加力和物体运动，例如滚动、下落和滑动。根据测试平台设计，我们评估智能体的两种能力：一种新颖性应用于不同物理场景时的表现，以及不同新颖性应用于同一物理场景时的表现。我们对人类玩家、学习型智能体和启发式智能体进行了全面评估。评估结果表明，人类的表现远超智能体。即使一些智能体在常规任务中表现良好，遇到新颖性时表现也会显著变差；能够适应新颖性的智能体，其适应速度通常也比人类慢。我们致力于推动智能体的发展，使其在开放世界物理环境中运行时能够达到或超越人类水平。

测试平台网站：https://github.com/phy-q/novphy

关键词：物理推理、开放世界学习、新颖性测试平台

1. 引言

人类智能的一个关键方面，是能够推理物体的物理行为，并在物理环境中做出决策[1]。研究表明，人类在出生后的第一年内就会发展出物理推理能力 [2, 3]。尽管开发出能够像人类一样进行通用物理推理的人工智能系统一直颇具挑战 [4, 5], ，但已有研究表明，人工智能系统在某些物理推理任务中的表现可以达到人类水平 [6, 7]。然而，问题在于：仅具备物理推理能力，是否足以在真实物理世界中顺利行动?

遇到新颖情境是现实世界的固有特征（图1）。作为人类，我们善于应对日常生活中不断遇到的这类新颖情境。例如，一个人能够在普通道路上骑自行车。下雨天路面湿滑时，此人仍能通过调整速度并平稳刹车，安全骑行而不打滑。尽管人类有时在面对新颖情境时，起初可能会做出次优决策，但在新颖性的影响下继续行动一小段时间后，通常能够迅速恢复并成功应对。同样，对于在开放世界物理环境中运行的智能体而言，除了具备物理推理能力，还必须具备应对新颖情境所需的能力 [8], ，即新颖性检测和适应能力。

与人类的智能相比，当前的人工智能系统在面对训练阶段中未曾出现，或开发者未曾预料到的情况时，往往会遇到困难 [11]。这可能是因为，试图解决这一问题的研究领域——开放世界学习（OWL）——相对较新 [12, 13, 14]。此外，缺乏足够的测试平台来实验和评估此类人工智能系统，也阻碍了其发展。目前已有诸如大富翁 [14], Polycraft [15], Cartpole [16], 等框架，将新颖性视为一等公民，并促进智能体实验。尽管其中一些是基于物理的环境 [16], ，但没有一个专门关注将新颖性引入智能体在现实世界中会遇到的物理场景。此外，这也超出了它们的



图1：现实世界中可能遇到的新颖性示例。左：自助结账机于21世纪初开始出现在超市中；在此之前，顾客习惯于传统结账方式，因此使用自助结账机对顾客来说是一种新颖性 [9]。右：交通事故通常是自动驾驶汽车面临的新颖情境，因为此类事件在训练数据中很少见，而且每起事故在视觉上通常都独一无二 [10]。

范畴，即在存在新颖性的情况下评估智能体执行现实世界物理任务的能力。

为弥补上述不足，我们提出了一个新的以新颖性为中心的测试平台 NovPhy。在该平台中，智能体需要在存在新颖性的情况下应对现实世界物理场景。NovPhy 涵盖了应用于不同物理场景的各种新颖性。我们基于基于物理的电子游戏《愤怒的小鸟》实现了该测试平台，因为它具有逼真的物理效果，是引入基于物理的新颖性的多功能领域。此外，《愤怒的小鸟》在物理推理研究 [5, 17, 18] 和开放世界学习（OWL）研究 [19, 20, 21]中都很受欢迎。本文的主要贡献可概括如下：

- **NovPhy——用于物理环境中新颖性检测与适应的测试平台：**我们在 NovPhy 中考虑了五种常见的物理场景：施加单一力和多个力，以及物体的滚动、下落和滑动。我们开发了代表多样化新颖性空间的八种新颖性。通过将这八种新颖性分别应用于五种物理场景，我们设计了任务模板，从而得到 40 个新颖任务模板。任务模板通过改变物体位置等任务模板参数来生成相关任务。我们还创建了 40 个不含新颖性的相应常规任务模板，以支持我们的评估协议。此外，我们开发了一个任务变体生成器，能够从这些任务模板中生成无限数量的任务。

• **开放世界物理环境的智能体评估设置：**我们提出了一套全面的评估设置，用于评估人工智能系统在开放世界物理环境中的新颖性检测和适应能力。在该设置中，新颖性与物理场景相互正交，因此便于我们在两种设置下评估智能体：第一，将相同的新颖性分别应用于多个物理场景的任务；第二，将多种新颖性分别应用于同一物理场景的任务。前者用于评估智能体在不同物理推理任务中应对相同新颖性的能力，后者则用于评估智能体在不同新颖情境下执行相同物理推理任务的能力。

• **确立基线智能体的结果：**我们评估了 11 个基线智能体：三个基于启发式的智能体、两个标准在线学习智能体、两个标准离线智能体、三个自适应学习智能体，以及一个随机智能体。我们报告了它们的新颖性检测和新颖性适应性能。

• **确立人类基线表现：**为了展示人类能够适应我们的新颖性，我们在测试平台上进行了涉及人类玩家的实验。这些结果还表明，与人工智能智能体相比，人类能够更好、更快地检测和适应新颖性，因此可作为人工智能需要达到的性能里程碑。

2. 背景与相关工作

本节将讨论物理推理研究、新颖性理论以及现有的以新颖性为中心的基准和测试平台的背景与相关工作。

2.1. 物理推理研究

随着日常运作对自主人工智能系统的依赖日益增加，物理推理已成为人工智能研究的重要方面。目前有多种物理推理基准和测试平台，可帮助人工智能系统提升物理推理能力，从而在现实世界中安全地运行。

Physion、 [7], IntPhys [22],CLEVERER [23]等物理推理基准基于视频，而 COPHY [24] 是一个基于图像的物理推理基准。Physion 包含八种物理推理

场景，包括滚动、滑动和抛射运动，这些都是在物理世界中行动所需的重要能力 [7]。另一方面，IntPhys 关注婴儿期习得的物理推理能力，例如客体永久性、时空连续性和形状恒常性 [22]。CLEVERER 基准提供视频，并提出受人类因果判断理论启发的问题 [23]。COPHY 基准提供一系列基于物理场景的图像，并要求预测对初始图像进行修改后的结果。COPHY 旨在测试应用于物理世界的反事实推理[24]。

为使人工智能系统的物理推理能力更贴近现实，研究人员创建了基于动作的基准，要求智能体采取行动以实现目标。此类基准的例子包括 PHYRE [25] 和 Virtual Tools [6]。PHYRE 是一个由简单的基于二维物理的任务组成的基准，旨在促进能够跨任务泛化的高效模型的开发 [25]。Virtual Tools 游戏侧重于评估智能体能否选择合适的工具，并使用该工具采取正确行动来解决任务 [6]。

Phy-Q [5] 也是一个基于动作的物理推理测试平台，包含 15 种丰富多样的物理推理场景。Phy-Q 中的一些物理场景（施加单个或多个力、滚动、下落、滑动）灵感源自童年时期发展出的物理推理能力；另一些场景（适时、清理路径和操控）则是机器人在物理环境中安全工作所需克服的挑战。作为对前述基准的改进，Phy-Q 测试平台支持基于不同泛化水平（局部泛化和广泛泛化）的多种评估设置，并且已测定人类在该测试平台各场景中的表现。此外，该测试平台还允许我们计算物理推理商，以反映智能体的物理推理能力。

尽管前述所有基准都支持开发具有先进物理推理能力的人工智能系统，但它们都不关注新情境下的物理推理，而这正是现实世界中的智能体经常会遇到的情形。在我们的测试平台 Nov-Phy 中，我们将物理推理场景与新颖情境相结合，构建一个类似现实世界的环境，以评估智能体的性能。为引入新颖情境，我们选取了 Phy-Q 测试平台中的前五种物理推理场景：单一力、多个力、滚动、下落和滑动。

2.2. 新颖性理论

人工智能系统已在广泛的封闭世界领域中展现出超越人类的性能 [26, 27, 28]。然而，与封闭世界相比，在开放世界中，智能体可能会因遇到新颖情境而难以完成任务。尽管开发者能够预测某些新颖情境，但仍有一些情境无法预料，因此不可能将所有可能的新颖情境都整合进智能体模型。DARPA 已启动“开放世界新颖性的人工智能与学习科学”（SAIL-ON）计划，旨在研究并发展创建人工智能系统所需的基础科学原理、通用工程技术和算法，使其能在新情境出现时进行恰当的自适应[12, 29]。

研究人员从不同角度探讨了如何对新颖性进行形式化定义。有些研究人员有时将新颖情境称为异常或分布外数据 [30, 15]。SAIL-ON 计划将新颖性定义为违反智能体关于外部世界（包括其他智能体、环境及其交互）的模型中隐含或明确假设的情境 [31, 14]。另一种观点是，Langley 将新颖性描述为环境中元素的变换 [13]。此类变换包括时空变换、结构变换、过程变换和约束变换。此外，Molineaux 和 Dannenhauer [32]对不同的环境变换进行了形式化定义。另一方面，Boult 等人 [30], 在人工智能领域提出了新颖性统一框架。Boult 等人着重关注世界空间、观测空间和智能体状态，以形式化定义各种类型的新颖性。

SAIL-ON 计划中的一个工作组——新颖性工作组（参见 [31]参与者名单）也制定了一个新颖性层级，以便根据新颖性的属性对其进行分类。这种新颖性分类有助于开展扎实的新颖性评估，因为它有助于设计覆盖广阔新颖性空间的新颖性，也有助于识别智能体模型无法胜任的不同类别的新颖性。本文采用了 SAIL-ON 新颖性工作组制定的新颖性层级。该新颖性层级包括物体、智能体、动作、关系、交互、环境、目标和事件 [31]。表 1 描述了新颖性层级的各个层级，以及我们在 NovPhy 中设计的代表性新颖性。第 3 节将详细讨论我们的设计要求：如何在物理环境中设计新颖性和新颖任务，从而对智能体进行全面评估。

2.3. 新颖性中心领域

能够促进智能体评估并比较智能体性能的新颖性中心领域，是推动开放世界学习发展的关键因素。为促进开放世界学习，已有若干测试平台、框架和基准将新颖性视为首要因素。本节介绍开放世界学习领域相关研究，并说明 NovPhy 在其中的定位。

GNOME 是一种以新颖性为中心的模拟器工具，可用于在策略棋盘游戏等多智能体环境中开发和评估人工智能系统 [14]。GNOME 被应用于热门棋盘游戏大富翁，以注入上述新颖性层级中前三级的新颖性。GNOME 中的新颖性示例包括在游戏中增加更多骰子、打乱棋盘上格子的顺序等。

NovGrid [33]和 NovelGridworlds [34]是为网格环境开发的两个框架。NovGrid 是一个用于在 MiniGrid 环境中生成新颖性的工具包 [35]。NovGrid 通过增强现有物体的功能，并使智能体能够检测新颖性并自适应，从而扩展了 MiniGrid 环境。NovelGridworlds 实现了一个受 Minecraft [36] 启发的网格世界，用于研究新颖性。在 NovGrid 中，作者引入了基于新颖性层级前两级的新颖性；而在 NovelGridworlds 中，作者引入了基于新颖性层级前三级的新颖性。

NovelCraft [15] 是一个基于修改版 Minecraft、用于新颖性检测与适应的基准数据集。NovelCraft 是一个三维环境，智能体需要选择一系列动作，将可用资源制成弹跳杆 [37]。在这项工作中，作者基于新颖性层级的第一级——对象 [15]——建立了智能体在新颖性方面的性能。

上述新颖性领域均不以物理学为基础，而这是开发能够在现实世界中运行的智能体时需要考虑的一项重要特征。带新颖性的倒立摆 [16] 和 Science Birds Novelty [19]是为新颖性检测和自适应而开发的两个基于物理学的领域。在倒立摆任务中，智能体必须通过向前、向后、向左或向右推动小车来保持杆的平衡。在基于物理游戏《愤怒的小鸟》 [38], 的 Science Birds Novelty 中，智能体需要用弹弓发射小鸟来摧毁猪。NovPhy 基于 Science Birds 框架 [19] ，并引入了各种真实的物理推理任务，因为我们的重点是评估智能体在新颖性影响下的物理推理能力。如第 2.1 节所述，我们使用五种物理推理情

景，并将其与新颖性层级全部八个级别中的新颖性相结合。

3. 在开放世界物理环境中设计新颖任务与智能体评估

本节将讨论我们在测试平台中设计任务以及为这些任务设置智能体评估时所遵循的要求。我们将包含新颖性的任务称为新颖任务，而将不包含新颖性的任务（即常规环境中的任务）称为常规任务。在测试平台中，我们考虑一种受限环境设定：智能体的动作仅作用于执行该动作时的环境状态，并由该动作决定环境随后的状态（即，动作执行后，智能体无法控制环境的后续状态）。

3.1. 设计新颖任务

在物理环境中工作的智能体必须应对不同的物理场景，例如对物体施加力、将物体从一处移动到另一处、避开路径上的障碍物等。我们在智能体可能遇到的这些物理环境场景中引入新颖性。

设计新颖任务时，**我们确保智能体必须在新颖性产生的影响下完成任务**。换言之，在新颖任务中，不存在能够绕过新颖性影响来完成任务的解。例如，考虑这样一种保龄球游戏：玩家必须将球滚过表面，以击倒球瓶。假设新颖性出现时，球滚过的表面会变得湿滑。在这种场景中，玩家必须与新颖元素（表面）互动才能完成任务。因此，成功完成这项任务的唯一方法，就是自适应地将球滚过湿滑的表面。

为确保智能体在完成任务时必须经历新颖性，在设计新颖性时，我们会考虑解决物理场景的过程中涉及的物理交互。我们将这些物理交互分为三个阶段：初始阶段、中间阶段和最终阶段。我们只设计至少会影响这三个交互阶段之一的新颖性，以确保智能体必须应对新颖性带来的影响。初始阶段包括智能体的动作对物体造成的即时影响；中间阶段包括动作即时影响所带来的后果；最终阶段则包括

完成任务的交互。在所有阶段中，也会考虑参与这些物理交互的物体。

例如，考虑前面提到的保龄球游戏。在这一场景中，可能发生的物理交互包括：玩家投出球并赋予球一个初始速度，球在表面上滚动，以及球击中球瓶并将其撞倒。在此例中，初始交互阶段包括球以及玩家施加给球的速度。中间阶段包括球和表面，以及球的滚动运动。最终阶段包括球和球瓶，以及球与球瓶之间的碰撞。因此，在这里，可以将新颖性应用于球、表面、球瓶，或任何会影响球滚动以及球与球瓶碰撞的因素，从而确保智能体必须在新颖性的影响下进行投球。

3.2. 设计智能体性能评估

在开放世界学习（OWL）中，智能体评估用于衡量智能体的两种能力：新颖性检测能力和新颖性适应能力 [39, 40, 41]。新颖性检测评估用于判断智能体能否成功检测环境中的新颖性，而新颖性适应评估用于判断智能体能否在存在新颖性的情况下成功执行任务。相较于新颖性检测，通常更强调新颖性适应，因为对于实际在开放世界环境中运行的智能体而言，在新颖性影响下完成任务比仅仅检测出某事物具有新颖性更为重要。在本研究中，我们也优先评估智能体的新颖性适应性能。

开放世界学习（OWL）的标准智能体评估设置由一系列试验组成，每次试验都包含一系列常规任务，随后是一系列新颖任务 [39, 41, 42]。我们在 NovPhy 中沿用相同的设置。我们认为，要衡量智能体是否真正能够自适应新颖性，**从常规任务转向新颖任务时，任务的解题路径应当发生变化**。从抽象层面来说，我们将解题路径定义为：由智能体的某个动作触发、包含相关物体的一系列物理交互，并最终促成任务解决。在为某个物理场景设计新颖任务时，我们也会设计一个对应的常规任务，其解题路径与新颖任务不同。随后，在为评估设定试验时，我们会选取这些常规任务与新颖任务的配对，以确保从常规任务转向新颖任务时，解题路径会发生变化。

在这种设定中，由于从常规任务到新颖任务的解题路径发生了明显变化，新颖性检测变得轻而易举。为了

检测是否存在新颖性，智能体只需监控常规任务中的解是否不再有效。为避免这一结果，可以单独设置一些试验，仅用于评估新颖性检测性能，其中常规任务和新颖任务采用相同的解题路径。在本研究中，我们的评估未纳入此类试验，因为我们的重点是评估智能体的新颖性适应性能。

在评估智能体在开放世界物理环境中的性能时，我们还考虑了另一项期望。从开放世界学习的角度来看，我们认为，**如果智能体在面对某种新颖性时确实能够执行任务，那么当这种新颖性应用于不同的物理场景时，该智能体也应能够在这些场景中执行任务**。此外，从物理推理的角度来看，我们认为，**如果智能体能够稳健地在某个物理场景中执行任务，那么当该场景受到不同新颖性的影响时，该智能体也应能够在该场景中执行任务**。为实现这两种评估设置，我们对新颖性和物理场景进行了正交设计，使同一种新颖性能够应用于多个场景，而同一个场景也能受到多种新颖性的影响。

4. NovPhy 测试平台

本节将介绍我们的测试平台、我们考虑的物理场景、设计的新颖性、测试平台中的任务，并说明我们在测试平台中采用的评估设置。

*4.1. NovPhy*简介

基于不同的物理场景和新颖性，我们利用《愤怒的小鸟》开发了以新颖性为中心的测试平台 NovPhy。我们使用了一款基于 Unity 开发的开源研究版游戏复刻作品，名为 Science Birds [43]。我们的测试平台改编自一个框架，该框架由 Science Birds [19] 开发而来，可用于注入新颖性并开展智能体评估。

在《愤怒的小鸟》中，玩家的目标是用弹弓发射一定数量的小鸟，消灭游戏关卡中的所有猪。在常规游戏环境中，除了弹弓，玩家还会遇到四种游戏物体：小鸟、猪、方块和平台。此外，我们还在常规环境中引入了一个名为“空气湍流”的外部智能体，它会对任何穿过其中的物体施加向上的作用力。外部智能体是具有目标导向行为的智能体；引入外部智能体后，可以出现阻碍或助力玩家行动的情况，而玩家采取的行动

在游戏中，小鸟、猪和方块是动态物体，其行为遵循牛顿物理学；平台则是静态的，不受外力影响。动态物体具有生命值，在发生碰撞时生命值会降低，生命值降至零时物体便会被摧毁。方块有 12 种形状变化，由以下 3 种材料之一制成：木材、石头和冰。猪分为三种类型，大小各不相同；体型越大，生命值越高。NovPhy 中的所有物体均见附录图 A.10。

游戏中的智能体可以随时请求当前游戏状态，形式为屏幕截图和/或符号表示。屏幕截图是游戏的 480x640 彩色图像。符号表示采用 JSON 格式，包含屏幕截图中的所有物体。在这里，每个物体都表示为一个由有序顶点构成的多边形，并附有其 8 位量化颜色的百分比。系统不会向智能体提供完整的世界状态，例如物体的确切位置及其质量、摩擦系数等物理参数，因为这些信息在现实世界中无法直接观察到。智能体的动作是小鸟相对于弹弓的释放点。有时，当游戏关卡中有不止一只小鸟时，智能体会依次执行一系列动作。我们提供了一个轨迹规划器，在给定目标点时，可在环境的默认设置下计算小鸟击中目标的释放点。如果智能体用指定数量的小鸟摧毁了所有猪，就通过该游戏关卡；否则即为失败。

本测试平台所考虑的物理场景和新颖性将在下文的 4.2 和 4.3 小节中分别讨论。

4.2. NovPhy 中的物理场景

如第 2 节所述，我们采用 Phy-Q 中介绍的前五种物理场景，因为它们是物理环境中最基本且最常见的场景。这些场景包括直接对目标物体施加力——单一力的作用和多重力的作用 [44]。与运动相关的场景包括滚动、下落和滑动，其灵感来源于人类婴儿期形成的物理推理能力[45]。以下列出了这五种场景，以及可用于实现相关任务目标的相应物理规则：

1. 单一力：必须通过单一力摧毁目标物体。
2. 多个力：必须使用多个力摧毁目标物体。

3. 滚动：必须让圆形物体沿着表面滚动至目标。4. 下落：物体必须落到目标上。5. 滑动：必须让非圆形物体沿着表面滑动至目标。

4.3. NovPhy 中使用的新颖性

我们针对 SAIL-ON 计划新颖性工作组提出的开放世界新颖性层级中的每个层级，设计了一种代表性新颖性。该新颖性层级由八个新颖性级别组成，涵盖了开放世界环境中可能出现的各种新颖性类型。表 1 展示了开放世界新颖性层级，以及 NovPhy 中代表性新颖性的描述。附录图 A.11 展示了为与游戏对象相关的新颖性引入游戏的新游戏对象。

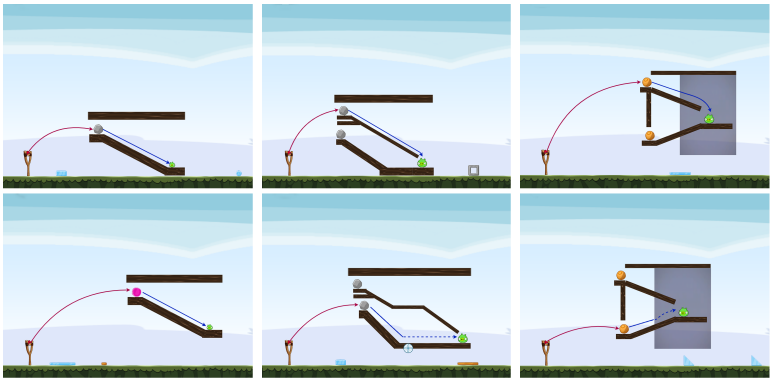
4.4. 任务模板

任务模板定义了一组相关任务，这些任务可以通过改变任务模板参数（例如物体的位置）来创建。我们将八种新颖性分别应用于上述章节讨论的五个物理场景，以设计任务模板。例如，对于滚动场景，我们设计了以下模板：要求在物体为新物体、受到风扇影响、空气湍流智能体改变向上作用力的大小、弹弓位于物体右侧、受到磁场影响、重力反转、空气湍流将物体向下推动，以及发射第一只小鸟后出现风暴时滚动物体。我们将应用了新颖性的物理场景称为新颖性场景。由于我们考虑了八种新颖性和五个场景，因此共有 40 个新颖性场景。

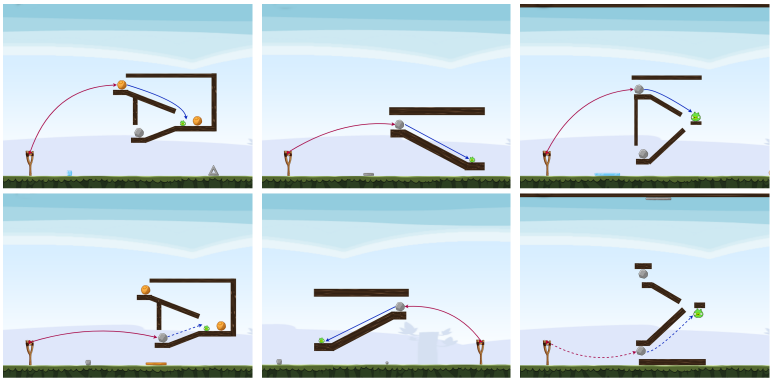
在为 NovPhy 设计新颖任务模板时，我们遵循第 3 节中讨论的要求。针对每个新颖性场景，我们都设计并手工制作了一个新颖任务模板。新颖任务模板是包含新颖性的任务模板。在设计过程中，我们确保新颖任务满足我们的要求，即必须在新颖性带来的影响下进行操作才能完成任务。附录表 C.4 详细说明了设计任务时如何满足这一要求，并考虑了引入新颖性后受到影响的解的物理交互阶段。随后，我们也为每个新颖任务模板设计了相应的常规任务模板。具体做法是移除新颖性

新颖性级别	描述	代表性新颖性
1. 物体	新的类别、属性、 or 表示 of 非自主实体。	一种新的猪/方块，具有 与普通猪不同的颜色 猪/方块。
2. 智能体	新的类别、属性， 或意志 实体的表示。	一种新颖的外部智能体——风扇， 可吹出空气（水平 from 从左向右），从而影响 物体的运动路径。
3. 动作	新的类别、属性， 或外部智能体行为的 新表示。	非新颖的外部智能体， 空气湍流，会增加 其向上作用力的大小。
4. 交互	新的类别、属性， 或动态属性的表示 会影响物体行 为的多个方面 个实体。	现有的圆形木质物体 现在具有磁性： 会排斥同类型的物体，并且 会吸引其他类型的物体。
5. 关系	新的类别、属性、 or 表示 of 静态属性 关系 之间 多个实体。	位于.....的弹弓 任务左侧的弹弓现在位于 任务的右侧（即， 空间关系 be- 弹弓与其他 物体之间发生了变化）。
6. 环境	新的类别、属性， 或元素的表示，独立于 特定实体。 特定实体。	环境中的重力 现在被反转，这 会影响 动态物体的行为。
7. 目标	新的类别、属性， 或表示形式，例如 部智能体的目标。	非新颖的外部智能体， 空气湍流，将其 目标从向上推物体 改为向下推物体。
8. 事件	新类别 ，属性 ， 或状态变 化序列的表示。	当第一只小鸟死亡时 ，a 发生了影响.....的风暴 物体的运动（通过向右施加 向右施加一个力，方- 方向）。

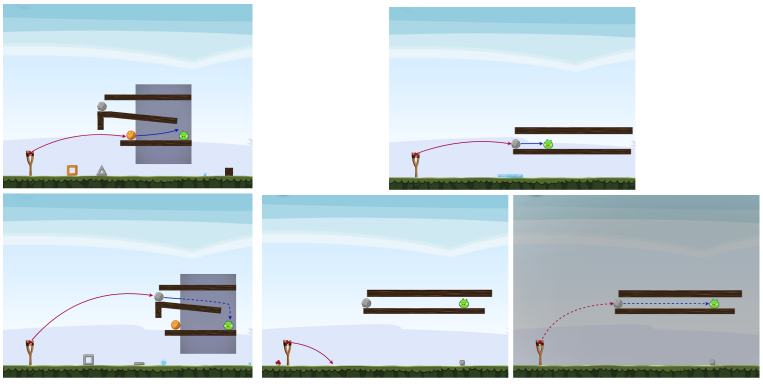
表 1：SAIL-ON 开放世界新颖性层级 [31] 以及 NovPhy 中各层级的代表性新颖性。



(a) 物体 (b) 智能体 (c) 动作



(d) 交互 (e) 关系 (f) 环境



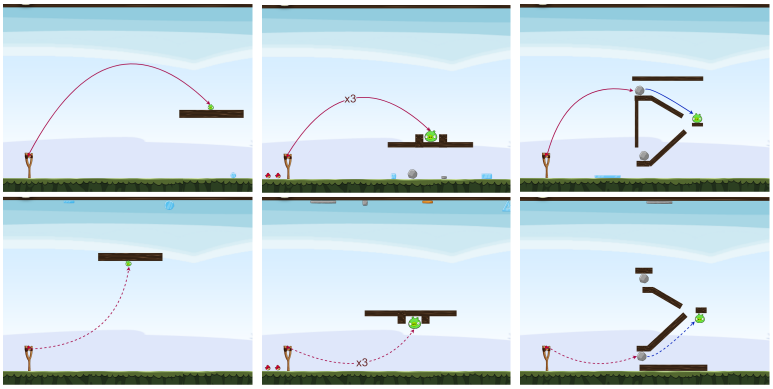
(g) 目标 (h) 事件

图2：滚动场景中八种新颖性的任务示例。在每个子图中，上方的图是常规任务，下方的图是带有相应新颖性的任务。箭头表示执行解时物体的运动轨迹。

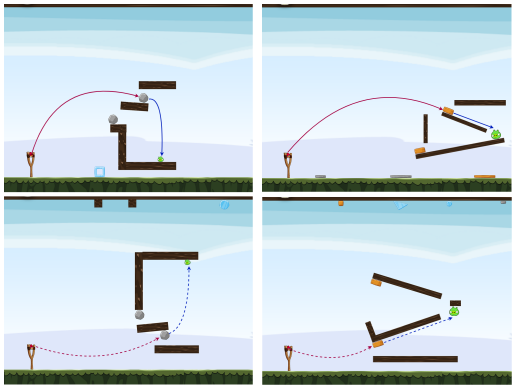
新颖任务模板，并相应调整该模板，使其在不包含新颖性的情况下仍可解。考虑到下一个期望，在设计常规任务模板时，我们还确保常规任务到新颖任务的解题路径会发生变化。

在任务模板设计中，我们还确保给定场景的所有模板都可以通过该场景相关的物理规则求解，该规则已在第4.2节中讨论。图2展示了如何将八种新颖性应用于滚动场景的任务。在每个子图中，上方为常规任务，下方为包含相应新颖性的任务。箭头表示执行解时物体的轨迹（红色表示小鸟的轨迹，蓝色表示小鸟受到撞击时其他物体的轨迹）。虚线箭头表示受新颖性影响的轨迹。要解决每个子图所示的新颖任务：

(a) 物体：新的粉色物体必须滚动。(b) 智能体：必须利用新风扇作用体的力，使物体继续滚动。(c) 动作：增大空气湍流智能体向上作用力的大小，从而帮助球沿斜坡向上滚动。(d) 交互：圆形木质（棕色）物体具有磁性，会排斥同类物体并吸引石头（灰色）物体，帮助石头沿斜坡向上滚动。(e) 关系：弹弓现在放置在任务右侧，因此不再从左向右发射，而是必须从右向左发射，使物体向左滚动。(f) 环境：反向重力使物体被吸引向天空，因此物体可以沿斜坡向上滚动。(g) 目标：目标发生了变化：空气湍流智能体从推动物体向上变为推动物体向下，这会阻碍物体在平坦表面上滚动，却有助于物体在倾斜表面上滚动。(h) 事件（如底部第一幅图所示，新颖模板中有两只小鸟；第一只小鸟死亡后，会触发如底部第二幅图所示的风暴）：第一只小鸟被消耗后，风暴便会发生，并对移动中的物体施加向右的力。因此，向圆形物体发射第二只小鸟，可使该物体继续滚动，最终到达猪所在的位置。



(a) 单一力　(b) 多个力　(c) 滚动



(d) 下落　(e) 滑动

图3：反向重力新颖性应用于五种物理场景的任务示例。在每种场景中，上图是常规任务，下图是新颖任务。箭头表示执行解时物体的轨迹。实线箭头表示不受新颖性影响的物体的轨迹，虚线箭头表示受新颖性影响的物体的轨迹。反向重力使所有动态物体都朝天空吸引，因此它们被平台阻挡，以避免飞走。动态物体的运动也受到反向重力的影响。

图3展示了反向重力这一新颖性如何应用于五种物理场景中的各项任务。NovPhy 中的全部 40 个任务模板见附录 B。

4.5. 任务生成

我们开发了一种任务生成器，可以根据给定的模板生成无限数量的任务。由任务模板生成的游戏关卡称为该模板的任务。在生成

任务时，我们会在关卡空间内的合适范围内改变游戏对象的位置。此外，还会在游戏关卡的随机位置添加一些随机游戏物体作为干扰对象，以迷惑智能体。在生成过程中，我们会确保仍然可以通过原始模板中的解题路径解决任务。为此，我们定义了模板特定约束，例如哪些游戏物体小鸟可以或无法触及、哪些物体应从干扰对象中排除、特定物体可以放置在哪些可行区域等。这些约束由模板设计者确定，并输入任务生成器。

我们为每种任务模板提供 350 个生成的任务，同时也提供任务生成器，以便在需要时生成更多任务。附录中的图 B.17 展示了应用反重力新颖性后，根据滚动场景模板生成的任务变化情况。

4.6. 评估协议

在 NovPhy 测试平台中，与标准的开放世界学习评估一样，我们评估智能体的新颖性检测和新颖性适应能力。在新颖性检测评估中，我们衡量智能体能否检测给定任务中是否存在新颖性。在新颖性适应评估中，我们衡量存在新颖性时智能体的任务性能。新颖性检测和新颖性适应评估均采用试验设置[39]。一次试验由一系列任务组成，先执行常规任务，经过随机数量的常规任务后切换到新颖任务。切换到新颖任务后，直到试验结束，后续所有任务都是新颖任务。图 4 展示了如何通过试验设置进行评估。试次集由一组试验构成。一个试次集包含同一新颖性场景下的试验。也就是说，给定试次集中的所有试验都只包含来自同一新颖性场景的常规任务和新颖任务。智能体不得在不同试验之间共享知识；也就是说，在某个新颖性场景的每次试验开始时，智能体都处于相同的初始状态，与评估开始时的状态一致。

要在新颖性场景上评估智能体，需要在该新颖性场景的常规任务分布上训练智能体，并使用该新颖性场景的试次集进行测试。这种评估类似于对智能体进行的局部泛化评估，即智能体在常规模板的任务上训练，并在同一模板的任务以及相应新颖模板的任务上进行测试。尽管该测试平台支持广泛泛化评估（即，智能体在物理场景的常规模板任务上训练。随后，在同一场景的不同模板任务及其对应的新颖模板任务上测试智能体），我们仍采用

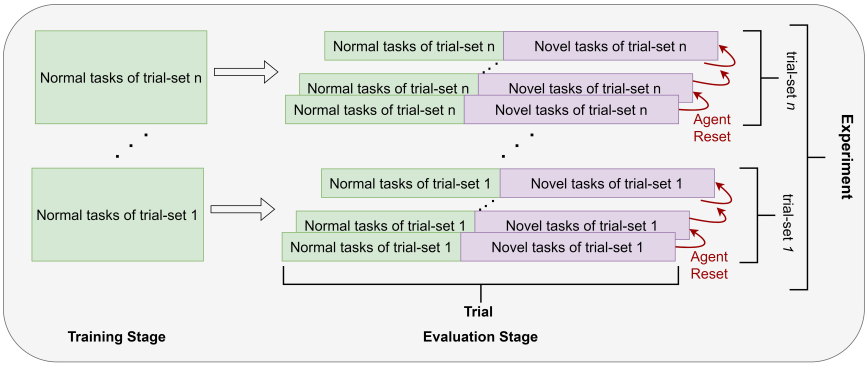


图 4：NovPhy 中使用的基于试次的评估协议。评估阶段遵循训练阶段。一次实验包含 *trial-sets*，其中一个试次集包含多个试验。一个试验包含数量可变的任务，这些任务首先从常规任务分布中抽取，然后从新颖任务分布中抽取。在给定的试次集中，只针对单一新颖性场景对智能体进行评估。

这种基于局部泛化的评估设置，因为已有研究证明，学习型智能体在物理推理中仍难以实现广泛泛化 [5]。此外，在适应新颖任务之前，智能体具备良好的常规任务表现是必要前提。

当智能体执行一个试验时，会记录任务完成状态（任务是否通过/失败）以及智能体在任务结束时获得的得分。这些数据用于计算智能体的新颖性适应性能。为了计算新颖性检测性能，智能体必须指出它认为新颖性出现于该试验中的哪个任务。

在本文中，我们重点在以下两种评估设置下评估智能体。

1. 新颖性知情评估：在此评估中，每次试验中新颖性出现时，都会通知智能体。仅评估智能体的新颖性适应能力。2. 新颖性非知情评估：在此评估中，将同时评估智能体的新颖性检测和新颖性适应能力。在给定试验中新颖性出现时，不会通知智能体。

4.7. 评估指标

对于新颖性检测，我们采用 SAIL- ON 项目中使用的标准开放世界学习指标：正确检测的试次（CDT）所占的百分比，以及

检测延迟（DD），通过检测到新颖性所需任务数的平均值计算得出 [39]。考虑一次试验 $t \in T$，其中 T 表示某个新颖性场景的一组试验， FP_t 表示试验 t 中智能体错误地将常规任务检测为新颖任务的数量。 TP_t表示试验 t 中智能体正确地将新颖任务检测为新颖任务的数量。正确检测的试验是指智能体仅在进入新颖任务序列后才检测到新颖性的试验。CDT 定义见公式 1。CDT 的取值范围为 0 到 1，1 为最佳结果。

$$CDT = \frac{1}{|T|} \sum_{t=1}^{|T|} \begin{cases} 1, & \text{if } FP_t = 0 \text{ and } TP_t \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

DD 通过正确检测新颖性所需的任务数量来量化检测延迟。DD 的定义见公式 2。DD 越低，检测性能（就及时性而言）越好；DD 的最优值为 1。

$$DD = \frac{1}{N_{cdt}} \sum_{t=1}^{|T|} \begin{cases} d_t, & \text{if } FP_t{=}0 \text{ and } TP_t \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

其中,

$$N_{cdt} = \sum_{t=1}^{|T|} \begin{cases} 1, & \text{if } FP_t{=}0 \text{ and } TP_t \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

而 d_t 是智能体在试验t 中报告检测到新颖性之前经历的新颖任务数量（包括智能体检测到新颖性的那个任务）。

为了衡量新颖性适应性能，我们使用通过率曲线（成功曲线）下的面积。首先，在给定的新颖性场景中，为了衡量智能体适应新颖性后的性能，我们使用渐近任务的通过率。我们将这一指标称为渐近性能（AP）。在渐近性能（AP）的公式 4 中， n表示新颖任务序列的长度， m 表示我们所考虑的渐近长度。渐近长度可根据试验中新颖任务所占的百分比进行调整。

$$AP = \frac{1}{m} \sum_{i=n-m}^{n} \left(\frac{1}{|T|} \sum_{t=1}^{|T|} \begin{cases} 1, & \text{if } i^{th} \text{ novel task in } t^{th} \text{ trial is passed} \\ 0, & \text{otherwise} \end{cases} \right) \tag{4}$$

其次，由于平均精度 无法体现适应的及时性，我们计算成功曲线下的总面积（AUS）。AUS 的定义如下。

$$AUS = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{|T|} \sum_{t=1}^{|T|} \begin{cases} 1, & \text{if } i^{th} \text{ novel task in } t^{th} \text{ trial is passed} \\ 0, & \text{otherwise} \end{cases} \right) \tag{5}$$

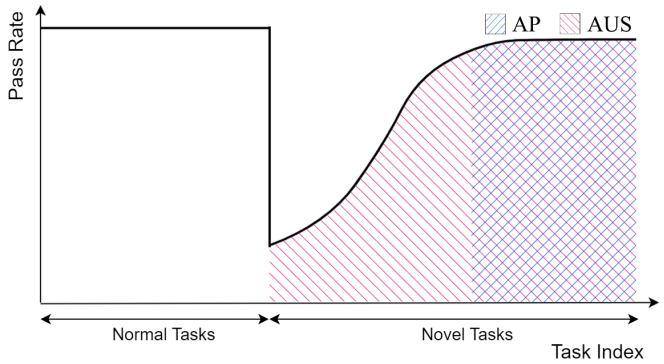


图 5：智能体完成一组试验后得到的通过率曲线示例。蓝色阴影部分用于计算渐近性能（AP），红色阴影部分用于计算成功曲线下面积性能（AUS）。

平均精度和AUS的取值范围均为0到1，其中1表示可达到的最佳适应性能。图5以图示方式展示了在计算平均精度和AUS时，智能体通过率曲线中所考虑的面积。

基于上述新颖性检测和新颖性适应指标，我们定义了 NPS（场景中的新颖性性能）和 SPN（新颖性下的场景性能）。NPS 衡量智能体在单一新颖性下、面对不同物理场景时的表现。SPN 衡量智能体在单一物理场景下、面对应用于该场景的不同新颖性时的表现。计算 NPS 指标时，对单一新颖性在所有场景中的平均性能取平均值。形式上，设 i 表示 i^{th} 新颖性，j 表示 j^{th} 场景。对于给定的新颖性 i，NPS 指标 NPS_{CDT}、$_i$、NPS_{DD}、$_i$、NPS_{AP}、$_i$ 和 NPS_{AUS}、$_i$ 分别定义为所有具有新颖性 i 的场景 j 中 CDT_{ij}、DD_{ij}、AP_{ij} 和 AUS_{ij} 的平均值。类似地，对于 SPN 指标，给定场景 j，对该场景下所有新颖性 i 的平均性能取平均值。也就是说，SPN 指标 SPN_{CDT}、$_j$、SPN_{DD}、$_j$、SPN_{AP}、$_j$ 和 SPN_{AUS}、$_j$，分别根据场景 j 中所有新颖性 i 对应的 CDT_{ij}、DD_{ij}、AP_{ij} 和 AUS_{ij} 的平均性能计算得出。

5. 实验

我们在 40 个新颖性场景中对基线智能体进行实验，以衡量智能体如何检测并自适应每个新颖性

场景。此外，我们还测定了人类在所有新颖性场景中的表现。本节介绍我们提供的基线智能体，以及各项实验所采用的实验设置。

5.1. 基线智能体

我们纳入了11个基线智能体的实验结果，这些智能体包括三个启发式智能体、七个学习型智能体和一个Random智能体。

启发式智能体。 启发式智能体基于由智能体开发者制定的硬编码物理规则。这些智能体均为参加AIBIRDS竞赛的智能体 [46]，；AIBIRDS竞赛是一项每年举办的竞赛，旨在寻找最优秀的《愤怒的小鸟》游戏AI智能体。以下是在NovPhy中评估的启发式智能体列表。

- Datalab：Datalab是一种具有六种策略的规划智能体。这些策略包括摧毁猪、摧毁物理结构以及射击圆形方块。该智能体根据可用的游戏对象、可能的轨迹、弹弓上的小鸟以及剩余的小鸟来选择要采取的行动 [47]。

- Eagle's Wing：Eagle's Wing智能体根据结构分析从五种策略中进行选择。这五种策略包括：射击未受保护的猪、尽可能摧毁更多方块，以及射击靠近圆形物体的物体 [48]。

- Pig Shooter：Pig Shooter 只有一种策略：射击猪。智能体会随机选择要射击的猪以及要使用的轨迹[49]。

所有这些启发式智能体都在无信息评估设置下运行，在该设置中，当试验中首次出现新颖任务时，不会告知智能体。

学习型智能体。 在这项工作中，我们评估了七种学习型智能体/智能体版本。本文介绍的所有七种学习型智能体均采用知情评估设置，即当新颖性在试验中出现时，智能体会收到通知。因此，我们不评估这些智能体的新颖性检测性能。这七种智能体分别是 *DQN* 离线/在线/自适应、关系离线/在线/自适应和朴素自适应。与 [5]，中使用的深度强化学习智能体相同，我们训练了一个 DQN [50, 51] 智能体，以及一个包含关系模块的关系型智能体 [52]。这两种智能体均在

由常规任务模板生成的任务上进行训练，并在包含相应新颖任务模板所生成任务的试验中进行评估。我们对 DQN 和关系型智能体的三个版本进行评估：离线、在线和自适应。在离线版本中，两个深度强化学习智能体 DQN 和关系型智能体在整个试验过程中始终选择 Q 值最高的动作。另一方面，在线学习智能体会在引入新颖性后更新 Q 网络，并尝试重新学习策略以解决新颖任务。我们还将近期开发的开放世界学习组件$NAPPING$ [53] 与 DQN 和关系型智能体结合进行评估。我们将这些智能体称为 DQN Adapt 和 Relational Adapt。

该朴素自适应 基于 [5], 中的$Pig\ Shooter$ 智能体构建，该智能体只会向猪发射。朴素自适应 在新颖性出现前的游戏任务中采用$Pig\ Shooter$ 的策略。智能体获知新颖性已经出现后，会搜索能够解出游戏关卡的组合（物体、轨迹和延迟），并记录每个尝试过的三元组。找到适用于某次试验的解三元组（例如，解三元组可以是（猪$_2$, 高轨迹，延迟 5 秒））后，朴素自适应 会持续使用该三元组，直到它无法再解出任务，此时智能体便会开始搜索另一个三元组。

随机智能体。 随机智能体会相对于弹弓随机选择一个释放点（x，y）。x 从 [-200, 200] 中采样，y 从 [-200, 200]中采样。该智能体适用于无信息评估设置。

5.2. 实验设置

5.2.1. 人类实验设置

涉及人类参与者的实验已获澳大利亚国立大学人类伦理委员会批准，审批协议为协议 2021/293。参与完全出于自愿，且未提供任何金钱补偿。共有 47 名参与者，年龄介于 20 至 35 岁，男女均有。他们都不是经验丰富的《愤怒的小鸟》玩家。部分参与者从未玩过这款游戏，另一些人了解基本的游戏机制，过去偶尔玩过，但对游戏策略并没有深入了解。参与者同意使用其游戏数据。

对于单个参与者，我们提供了来自 10 个新颖性场景的 10 次试验。我们设置了四个这样的试验集，以涵盖全部 40 个新颖性场景。在单次试验中，包含 1–4 个常规任务和 4 个新颖任务。参与者平均花费 25–30 分钟完成实验。参与者尝试解决

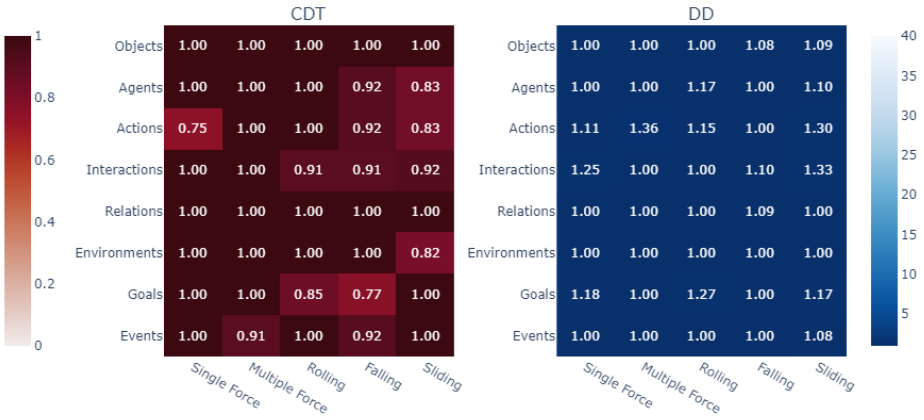


图 6：人类玩家的 CDT（左）和 DD（右）结果。热力图中，横轴表示物理场景，纵轴表示所应用的新颖性。

这些任务，并在每项任务结束时说明自己是否检测到了新颖性。

5.2.2. 智能体实验设置

我们对所有智能体均采用标准的 SAIL-ON 评估设置。如前所述，我们有八种新颖性和五种物理推理场景，共构成 40 个新颖性场景。对于每个新颖性场景，我们在 40 次试验中测试智能体。一次试验包含 1–40 个常规任务和 40 个新颖任务。所有启发式智能体和随机智能体都不需要训练。然而，学习型智能体会先在相应新颖性场景下的常规任务上进行训练，然后在试验集上接受评估。

6. 结果与分析

本节将呈现并讨论我们所开展实验的结果：人类玩家实验和基线智能体实验。对于这两项实验，我们均报告新颖性检测性能和新颖性适应性能。

6.1. 人类表现

图 6 展示了人类玩家的 CDT 和 DD 结果。总体而言，参与者几乎在所有试验中都能正确检测到新颖性（CDT 接近 1）。CDT 最低的是新颖性场景

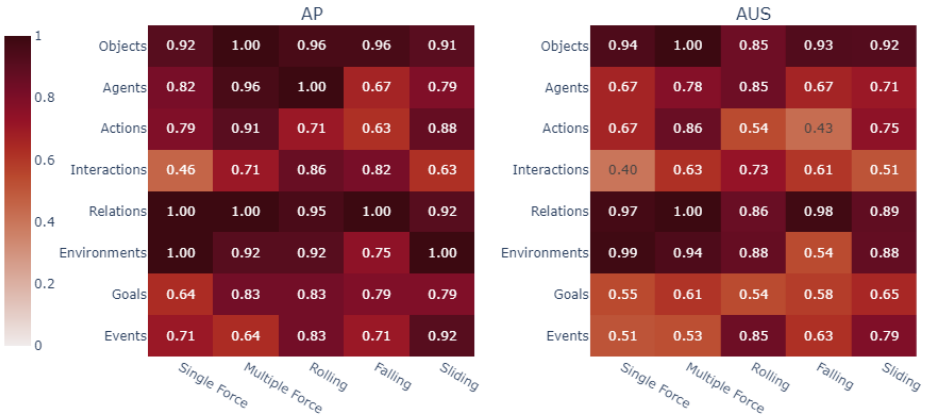


图 7：人类玩家的平均精度（左）和 AUS（右）结果。在热力图中，横轴为物理场景，纵轴为所应用的新颖性。计算平均精度时考虑的渐近长度为 2。

动作-单一力，其中空气湍流智能体的向上作用力有所增加。这可能是因为只有在交互之后，这种新颖性才会在视觉上显现。此外，将这种新颖性应用于单力场景时，与常规任务中的射击角度相比，玩家只需稍微调整小鸟的射击角度。由于这种新颖性的影响较为细微，人类可能无法察觉。CDT 第二低的新颖性场景——目标-下落——也可能是如此。观察 DD 结果可知，在大多数情况下，参与者能在首次遇到新颖性的游戏关卡中检测到它（DD 接近 1）。总体而言，对于那些在交互前无法通过视觉察觉的新颖性（动作、交互和目标），与其他可在交互前通过视觉察觉的新颖性相比，人类的检测延迟更长。

人类玩家的平均精度和 AUS 性能如图 7 所示。在平均精度计算中，我们将渐近长度设为 2（即试验中最后两个任务的性能）。如平均精度结果所示，参与者在大多数新颖性场景中都取得了 80% 以上的性能。总体而言，对于平均精度和 AUS，参与者在需要高度精准动作的新颖性场景中表现较差。例如，交互-单力、目标-单一力、事件-多力、交互-滑动和动作-下落在两项指标中的结果都略低，这是因为要解决

这些新颖性场景中的任务，需要以很高的准确度发射小鸟。适应速度通过 AUS 结果体现出来。当准确度要求更高时，人类需要更多任务才能适应新颖性。这反映在要求精准动作的新颖性（交互、目标和事件）相对较低的 AUS 结果上。另一方面，不要求精准动作的新颖性（例如物体和关系）的 AUS 结果几乎达到完美，因为人类能够立即适应这类新颖性。

此外，我们使用非参数曼-惠特尼检验和斯皮尔曼秩相关，分析了人类玩家的检测性能与适应性能之间的关系（详情见附录 F）。考察 CDT 与适应性能（包括平均精度和 AUS）之间的关系时，只有在动作-单一力新颖性场景中，结果表明适应性能与参与者是否检测到新颖性无关。正如上文所述，这是因为尽管人类已成功适应动作-单一力，但人类可能无法感知到其中的新颖性。检测延迟与适应性能（包括平均精度和 AUS）之间的相关性表明，一些新颖性场景（如目标-单一力、目标-滚动和交互-滑动）存在中等程度的负相关，这意味着玩家检测到新颖性所花的时间越长，适应性能就越低。相关性图见附录图 F.44。

6.2. 基线智能体性能

本节将讨论 11 个基线智能体和人类在新颖性检测与新颖性适应方面的性能。

6.2.1. 基线智能体新颖性检测性能

如第 5.1 节所述，NovPhy 中的一些智能体在无信息评估下运行。我们计算了这些智能体的检测性能指标。这些基线智能体中的检测模块基于通过率偏差（详见附录 D）。表 2 展示了所有启发式智能体、随机智能体和人类按新颖性划分的检测指标，表 3 则展示了按场景划分的检测指标。

结果显示，Datalab 的总体 CDT 最高，而 Pig Shooter 的最低。Pig Shooter 的 CDT 最低是意料之中的，因为该智能体只会直接射击猪，因此在同一试验的各项任务中通常会得到相同结果，所以通过率不会出现显著偏差。就总体 DD 而言，Random 智能体具有

Novelty	Measure	Human	DQN Offline	DQN Online	DQN Adapt	Rel Offline	Rel Online	Rel Adapt	Naive Adapt	Datalab	Eagle'sWing	PigShooter	Random
1. Objects	$NPS_{CDT,1}$	1.00 ± 0.00	-	-	-	-	-	-	-	0.65 ± 0.13	0.57 ± 0.08	0.25 ± 0.08	0.23 ± 0.09
	$NPS_{DD,1}$	1.03 ± 0.02	-	-	-	-	-	-	-	7.85 ± 4.04	5.46 ± 1.17	6.07 ± 2.53	12.79 ± 4.31
	$NPS_{AP,1}$	0.95 ± 0.02	0.65 ± 0.17	0.66 ± 0.14	0.79 ± 0.10	0.54 ± 0.10	0.50 ± 0.10	0.80 ± 0.06	0.61 ± 0.19	0.37 ± 0.19	0.40 ± 0.21	0.02 ± 0.02	0.01 ± 0.01
	$NPS_{AUS,1}$	0.93 ± 0.02	0.66 ± 0.17	0.67 ± 0.16	0.81 ± 0.10	0.56 ± 0.11	0.55 ± 0.10	0.80 ± 0.06	0.61 ± 0.19	0.38 ± 0.19	0.41 ± 0.21	0.02 ± 0.02	0.01 ± 0.01
2. Agents	$NPS_{CDT,2}$	0.95 ± 0.03	-	-	-	-	-	-	-	0.49 ± 0.17	0.42 ± 0.12	0.17 ± 0.07	0.24 ± 0.07
	$NPS_{DD,2}$	1.05 ± 0.03	-	-	-	-	-	-	-	10.96 ± 4.29	10.01 ± 2.25	4.57 ± 0.39	19.59 ± 1.42
	$NPS_{AP,2}$	0.85 ± 0.02	0.06 ± 0.03	0.21 ± 0.07	0.73 ± 0.13	0.07 ± 0.04	0.11 ± 0.04	0.67 ± 0.09	0.11 ± 0.06	0.05 ± 0.03	0.05 ± 0.03	0.01 ± 0.01	0.01 ± 0.00
	$NPS_{AUS,2}$	0.73 ± 0.03	0.06 ± 0.03	0.18 ± 0.06	0.62 ± 0.10	0.07 ± 0.04	0.10 ± 0.04	0.56 ± 0.09	0.09 ± 0.05	0.05 ± 0.03	0.05 ± 0.03	0.00 ± 0.00	0.01 ± 0.00
3. Actions	$NPS_{CDT,3}$	0.90 ± 0.05	-	-	-	-	-	-	-	0.44 ± 0.15	0.49 ± 0.19	0.24 ± 0.02	0.17 ± 0.06
	$NPS_{DD,3}$	1.19 ± 0.07	-	-	-	-	-	-	-	10.11 ± 1.69	8.12 ± 2.20	5.67 ± 0.57	19.27 ± 1.73
	$NPS_{AP,3}$	0.78 ± 0.05	0.07 ± 0.06	0.13 ± 0.08	0.56 ± 0.17	0.05 ± 0.03	0.12 ± 0.07	0.63 ± 0.16	0.10 ± 0.07	0.03 ± 0.02	0.08 ± 0.04	0.00 ± 0.00	0.01 ± 0.00
	$NPS_{AUS,3}$	0.65 ± 0.08	0.08 ± 0.06	0.13 ± 0.08	0.53 ± 0.15	0.05 ± 0.03	0.10 ± 0.06	0.55 ± 0.14	0.08 ± 0.05	0.03 ± 0.02	0.08 ± 0.04	0.00 ± 0.00	0.01 ± 0.00
4. Interactions	$NPS_{CDT,4}$	0.95 ± 0.02	-	-	-	-	-	-	-	0.42 ± 0.15	0.46 ± 0.18	0.18 ± 0.05	0.55 ± 0.12
	$NPS_{DD,4}$	1.14 ± 0.07	-	-	-	-	-	-	-	5.87 ± 1.32	16.31 ± 7.38	5.25 ± 0.35	18.40 ± 5.13
	$NPS_{AP,4}$	0.69 ± 0.07	0.10 ± 0.04	0.24 ± 0.07	0.75 ± 0.16	0.08 ± 0.03	0.19 ± 0.05	0.72 ± 0.15	0.33 ± 0.14	0.10 ± 0.06	0.3 ± 0.17	0.03 ± 0.02	0.06 ± 0.02
	$NPS_{AUS,4}$	0.58 ± 0.05	0.10 ± 0.04	0.25 ± 0.08	0.64 ± 0.14	0.08 ± 0.03	0.19 ± 0.04	0.63 ± 0.13	0.27 ± 0.11	0.18 ± 0.07	0.35 ± 0.18	0.02 ± 0.02	0.06 ± 0.02
5. Relations	$NPS_{CDT,5}$	1.00 ± 0.00	-	-	-	-	-	-	-	0.27 ± 0.18	0.28 ± 0.17	0.08 ± 0.07	0.01 ± 0.01
	$NPS_{DD,5}$	1.02 ± 0.02	-	-	-	-	-	-	-	3.47 ± 0.40	7.49 ± 0.02	3.86 ± 0.54	4.00 ± 0.00
	$NPS_{AP,5}$	0.97 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	$NPS_{AUS,5}$	0.94 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
6. Environments	$NPS_{CDT,6}$	0.96 ± 0.04	-	-	-	-	-	-	-	0.44 ± 0.15	0.35 ± 0.13	0.01 ± 0.01	0.40 ± 0.10
	$NPS_{DD,6}$	1.00 ± 0.00	-	-	-	-	-	-	-	4.75 ± 0.19	5.50 ± 1.03	2.50 ± 0.95	14.64 ± 0.83
	$NPS_{AP,6}$	0.92 ± 0.05	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.03 ± 0.01
	$NPS_{AUS,6}$	0.85 ± 0.08	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.03 ± 0.01
7. Goals	$NPS_{CDT,7}$	0.92 ± 0.05	-	-	-	-	-	-	-	0.43 ± 0.13	0.55 ± 0.10	0.19 ± 0.05	0.39 ± 0.09
	$NPS_{DD,7}$	1.12 ± 0.05	-	-	-	-	-	-	-	11.62 ± 1.48	12.28 ± 2.99	5.73 ± 0.22	15.06 ± 2.40
	$NPS_{AP,7}$	0.78 ± 0.04	0.07 ± 0.02	0.09 ± 0.03	0.45 ± 0.15	0.04 ± 0.02	0.14 ± 0.08	0.54 ± 0.18	0.29 ± 0.12	0.08 ± 0.04	0.09 ± 0.03	0.01 ± 0.01	0.02 ± 0.01
	$NPS_{AUS,7}$	0.59 ± 0.02	0.08 ± 0.03	0.10 ± 0.03	0.42 ± 0.13	0.04 ± 0.02	0.12 ± 0.05	0.49 ± 0.16	0.29 ± 0.12	0.08 ± 0.04	0.09 ± 0.04	0.01 ± 0.01	0.02 ± 0.01
8. Events	$NPS_{CDT,8}$	0.97 ± 0.02	-	-	-	-	-	-	-	0.46 ± 0.13	0.46 ± 0.12	0.07 ± 0.03	0.26 ± 0.80
	$NPS_{DD,8}$	1.02 ± 0.02	-	-	-	-	-	-	-	7.36 ± 1.73	11.00 ± 2.36	4.80 ± 1.35	18.14 ± 2.18
	$NPS_{AP,8}$	0.76 ± 0.05	0.27 ± 0.12	0.21 ± 0.10	0.21 ± 0.1	0.29 ± 0.11	0.2 ± 0.09	0.2 ± 0.09	0.00 ± 0.00	0.13 ± 0.05	0.12 ± 0.04	0.01 ± 0.00	0.02 ± 0.01
	$NPS_{AUS,8}$	0.66 ± 0.07	0.27 ± 0.12	0.24 ± 0.11	0.23 ± 0.1	0.29 ± 0.11	0.22 ± 0.10	0.21 ± 0.09	0.00 ± 0.00	0.13 ± 0.05	0.12 ± 0.04	0.01 ± 0.00	0.02 ± 0.01

Table 2: Results of the NPS measures of the agents and humans for the eight novelties.

Scenario	Measure	Human	DQN Offline	DQN Online	DQN Adapt	Rel Offline	Rel Online	Rel Adapt	Naive Adapt	Datalab	Eagle'sWing	PigShooter	Random
1. Single Force	$SPN_{CDT,1}$	0.97 ± 0.03	-	-	-	-	-	-	-	0.59 ± 0.07	0.67 ± 0.05	0.08 ± 0.04	0.31 ± 0.10
	$SPN_{DD,1}$	1.07 ± 0.04	-	-	-	-	-	-	-	6.61 ± 0.88	7.42 ± 1.46	4.82 ± 1.02	14.21 ± 2.62
	$SPN_{AP,1}$	0.79 ± 0.07	0.11 ± 0.05	0.13 ± 0.05	0.37 ± 0.14	0.12 ± 0.06	0.13 ± 0.05	0.36 ± 0.13	0.13 ± 0.11	0.08 ± 0.04	0.16 ± 0.11	0.01 ± 0.00	0.03 ± 0.02
	$SPN_{AUS,1}$	0.71 ± 0.08	0.11 ± 0.05	0.13 ± 0.05	0.33 ± 0.12	0.12 ± 0.06	0.13 ± 0.06	0.33 ± 0.11	0.13 ± 0.12	0.08 ± 0.05	0.17 ± 0.12	0.01 ± 0.00	0.03 ± 0.02
2. Multiple Force	$SPN_{CDT,2}$	0.99 ± 0.01	-	-	-	-	-	-	-	0.55 ± 0.14	0.55 ± 0.10	0.14 ± 0.03	0.14 ± 0.09
	$SPN_{DD,2}$	1.05 ± 0.05	-	-	-	-	-	-	-	4.29 ± 1.28	4.88 ± 0.47	5.16 ± 0.38	17.48 ± 1.26
	$SPN_{AP,2}$	0.87 ± 0.05	0.07 ± 0.03	0.15 ± 0.06	0.11 ± 0.05	0.05 ± 0.03	0.06 ± 0.03	0.16 ± 0.08	0.17 ± 0.12	0.01 ± 0	0.08 ± 0.05	0.00 ± 0.00	0.01 ± 0.00
	$SPN_{AUS,2}$	0.79 ± 0.07	0.08 ± 0.03	0.14 ± 0.06	0.12 ± 0.06	0.04 ± 0.02	0.07 ± 0.03	0.15 ± 0.07	0.17 ± 0.12	0.05 ± 0.05	0.10 ± 0.07	0.00 ± 0.00	0.01 ± 0.00
3. Rolling	$SPN_{CDT,3}$	0.97 ± 0.02	-	-	-	-	-	-	-	0.26 ± 0.07	0.24 ± 0.08	0.20 ± 0.04	0.34 ± 0.07
	$SPN_{DD,3}$	1.07 ± 0.04	-	-	-	-	-	-	-	10.49 ± 3.11	12.02 ± 2.28	4.45 ± 0.42	16.13 ± 1.71
	$SPN_{AP,3}$	0.88 ± 0.03	0.24 ± 0.13	0.27 ± 0.12	0.59 ± 0.14	0.24 ± 0.11	0.22 ± 0.10	0.60 ± 0.14	0.18 ± 0.10	0.07 ± 0.03	0.07 ± 0.03	0.00 ± 0.00	0.02 ± 0.00
	$SPN_{AUS,3}$	0.76 ± 0.05	0.25 ± 0.13	0.28 ± 0.12	0.57 ± 0.13	0.25 ± 0.11	0.23 ± 0.11	0.57 ± 0.13	0.15 ± 0.08	0.07 ± 0.02	0.07 ± 0.03	0.00 ± 0.00	0.02 ± 0.00
4. Falling	$SPN_{CDT,4}$	0.93 ± 0.03	-	-	-	-	-	-	-	0.48 ± 0.11	0.44 ± 0.11	0.19 ± 0.05	0.36 ± 0.08
	$SPN_{DD,4}$	1.03 ± 0.02	-	-	-	-	-	-	-	10.04 ± 1.75	12.51 ± 2.37	4.89 ± 0.48	13.35 ± 1.98
	$SPN_{AP,4}$	0.79 ± 0.05	0.17 ± 0.11	0.20 ± 0.1	0.57 ± 0.14	0.14 ± 0.09	0.2 ± 0.08	0.55 ± 0.14	0.21 ± 0.09	0.19 ± 0.10	0.19 ± 0.10	0.02 ± 0.01	0.02 ± 0.01
	$SPN_{AUS,4}$	0.67 ± 0.07	0.18 ± 0.11	0.22 ± 0.10	0.51 ± 0.13	0.14 ± 0.09	0.19 ± 0.08	0.50 ± 0.13	0.19 ± 0.09	0.20 ± 0.10	0.19 ± 0.10	0.02 ± 0.01	0.02 ± 0.01
5. Sliding	$SPN_{CDT,5}$	0.93 ± 0.03	-	-	-	-	-	-	-	0.37 ± 0.13	0.33 ± 0.13	0.12 ± 0.06	0.25 ± 0.05
	$SPN_{DD,5}$	1.13 ± 0.04	-	-	-	-	-	-	-	9.10 ± 1.72	12.97 ± 4.41	6.99 ± 2.25	22.12 ± 2.83
	$SPN_{AP,5}$	0.85 ± 0.04	0.17 ± 0.11	0.21 ± 0.12	0.55 ± 0.15	0.13 ± 0.08	0.18 ± 0.07	0.55 ± 0.14	0.21 ± 0.10	0.13 ± 0.10	0.16 ± 0.11	0.02 ± 0.01	0.01 ± 0.00
	$SPN_{AUS,5}$	0.76 ± 0.05	0.18 ± 0.12	0.21 ± 0.12	0.49 ± 0.13	0.13 ± 0.08	0.18 ± 0.07	0.48 ± 0.12	0.19 ± 0.10	0.14 ± 0.10	0.16 ± 0.12	0.01 ± 0.01	0.01 ± 0.00

Table 3: Results of the SPN measures of the agents and humans for the five scenarios.



图8：智能体在不同新颖性下的通过率。横轴表示试验中任务的索引。索引 -40 至 -1 表示常规任务，0 至 40 表示新颖任务。纵轴表示相关新颖性下所有试验的平均通过率。

DD 最高。这是因为智能体动作中的随机性会导致它在试验中的随机位置检测到新颖性，使得整体 DD 接近该试验中任务数量的平均值。Pig Shooter 的 DD 最佳（最低），这是因为在少数能够正确检测新颖性的试验中，它都能迅速完成检测，这归功于其动作选择的确定性。所有智能体的新颖性检测性能都远不及人类，后者的表现近乎完美（CDT $= 0.96$ 和 DD $= 1.07$）。

6.2.2. 基线智能体新颖性适应性能

按新颖性划分的适应图见图8，按场景划分的适应图见图9。根据适应曲线得出的新颖性适应指标、按场景统计的平均精度（AP）和 AUS 结果列于表 2，按新颖性统计的结果列于表 3。平均精度（AP）结果基于新颖任务最后 50% 的数据（m=20）。理想情况下，智能体在常规任务中的通过率应较高；当新颖任务开始时（图8和图9中的索引 0），性能会下降，并



图9：各场景中智能体的通过率。横轴表示试验中任务的索引。索引 -40 至 -1 表示常规任务，0 至 40 表示新颖任务。纵轴显示相应场景中所有相关试验的平均通过率。

在几个任务内恢复其性能，达到常规任务表现。因此，理想智能体的平均精度（AP）和 AUS 结果应接近 1。

图9中，与其他智能体相比，DQN Adapt 和 Relational Adapt 在五种场景中的四种场景下表现出更好的适应行为，唯一的例外是多个力。在这四种场景中，比较这两个智能体可以发现，在考察 $SPN_{AP,j}$ 和$SPN_{AUS,j}$ 时，DQN Adapt 的性能略优于 Relational Adapt（显著性水平为 10%）。如图8所示，这两个智能体能够适应五种新颖性，但 Relations、Environments 和 Events 除外。这是因为，要适应 Relations 和 Environments 新颖性，智能体需要调整预定义的轨迹规划器。例如，在 Environments 新颖性中，轨迹是上下颠倒的。然而，-Adapt 智能体目前使用的是给定的轨迹规划器。同样，除单一力场景外，Naive Adapt智能体在所有场景中也表现出适应行为。该智能体在除 Relations、Environments 和 Events 之外的所有新颖性中都表现出适应行为。然而，由于 Naive Adapt智能体在一次试验中只搜索一个解元组，因此无法适应那些在不同情况下需要不同解元组的新颖性试验。因此，Naive Adapt智能体的性能未能达到 DQN Adapt 和 Relational Adapt 这两个智能体的水平；后两者能够通过每次试验高效地学习如何应对不同场景中的新颖性。总体而言，

这些智能体在物体新颖性方面表现出最佳的适应性能。这是因为此处测试的物体新颖性是现有物体颜色的变化，并未对智能体的动作造成重大影响。值得注意的是，在任何场景或任何新颖性中，都没有智能体达到人类的通过率；唯一的例外是交互新颖性，在该项中 Relational Adapt 和 DQN Adapt 超过了人类的 $NPS_{AP,j}$ 和$NPS_{AUS,j}$。然而，在与人类相同数量的任务中，这两个智能体都未能达到人类能够实现的性能，这表明适应效率仍有提升空间。

7. 结论与未来工作

NovPhy 的目标是促进人工智能系统的发展，使其能够在存在新颖性的情况下执行物理推理任务；这是开放世界物理环境中的系统可能会遇到的情形。为实现这一目标，NovPhy 旨在评估智能体在物理环境中检测新颖性并适应这些新颖性影响下的任务执行能力。我们针对现实世界中五种常见的物理场景设计了任务模板。随后，我们通过向这些任务模板引入多样的新颖性，设计了新颖任务。这种设计能够从两个方向衡量智能体的新颖性检测与适应能力：1）当新颖性应用于不同物理场景时，智能体在该新颖性下的表现；2）当不同新颖性应用于某一物理场景时，智能体在该物理场景中的表现。为衡量智能体真正的新颖性适应性能，在设计任务时，我们确保智能体必须在新颖性的影响下完成任务，而不能绕过新颖性来解决任务。我们采用试验设置对智能体进行评估：在该设置中，智能体必须先完成一个场景中不含新颖性的任务序列，随后完成同一场景中含有新颖性的任务序列。

我们使用人类玩家、学习型智能体和启发式智能体，建立了该测试平台的基线结果。结果表明，新颖性会严重影响智能体的性能，而且有些智能体随着完成的任务越来越多，能够逐渐恢复。不过，与人类相比，智能体的解题率和适应效率都不尽如人意。虽然我们的结果表明，DQN Adapt 和 Relational Adapt 智能体能够适应大多数新颖性，但仍有一些新颖性是智能体无法适应的。主要原因是，智能体仍然使用提供的轨迹规划器来规划射击（释放点）。未来关于智能体的研究可以

着眼于减少对轨迹规划器的依赖，从而使智能体能够适应更广泛的新颖性。

我们设想了 NovPhy 在多个方向上的改进空间。通过引入更多体现新颖性层级各层次的新颖性，可以进一步完善 NovPhy。此外，当智能体在现有场景中展现出高效的新颖性检测与新颖性适应能力后，测试平台还可加入更复杂的物理推理场景，例如相对高度、相对重量和清理路径。此外，测试平台还可扩展为评估智能体的新颖性表征能力（即评估智能体能否正确检测任务中的“新颖之处”）。此外，NovPhy 中在新颖性影响下评估物理推理能力这一理念，也可扩展至其他物理推理领域。例如，可以在 CausalWorld [54]和 RLBench [55], 等基于物理的机器人基准测试中引入新颖性，从而评估智能体在 Angry Birds 领域中未曾出现过的物理场景下的表现，例如拉动、拾取与放置以及堆叠。我们相信，NovPhy 为未来研究奠定了基础，有助于开发能够像人类一样在物理世界中高效检测并适应新颖性的智能体。

致谢

本研究由国防高级研究计划局（DARPA）和陆军研究办公室（ARO）资助，并依据合作协议编号 W911NF-20-2-0002 开展。本文件所载观点和结论仅代表作者观点，不应被解读为代表 DARPA、ARO 或美国政府明示或暗示的官方政策。无论本文是否注明版权，美国政府均有权为政府用途复制和分发再版文件。

参考文献

[1] E.Davis, 《物理推理》（2006）。网址：https://cs.nyu.edu/~davise/papers/handbookKR.pdf[2]R. Baillargeon, 《婴儿期的物理推理》, 《认知神经科学》（1995）181–204。[3] R. Baillargeon、J. Li、W. Ng、S. Yuan, 《关于婴儿的物理推理的解释》, 《学习与婴儿心智》66（2009）116。

[4] F. Chollet, 《论智能的衡量》（2019）。arXiv:1911.01547。[5] C. Xue*、V. Pinto*、C. Gamage*、E. Nikonova、P. Zhang、J. Renz, 《Phy-Q 作为物理推理智能的衡量标准》，《自然·机器智能》5（2023）83–93，*贡献相同。[6]K. R. Allen、K. A. Smith、J. B. Tenenbaum, 《通过模拟进行快速试错学习支持灵活的工具使用和物理推理》，《美国国家科学院院刊》117（47）（2020）29302–29310。doi:10.1073/pnas.1912341117。网址：https://www.pnas.org/content/117/47/29302.full.pdf[7] D. Bear、E. Wang、D. Mrowca、F. J. Binder、H.-Y. Tung、R. Pramod、C. Holdaway、S. Tao、K. A. Smith、F.-Y. Sun 等, 《Physion：评估人类和机器通过视觉进行物理预测的能力》，载于：第三十五届神经信息处理系统会议数据集与基准赛道（第一轮），2021年。[8] Z. Zeng、E. Davis, 《开放世界中的物理推理》，《认知系统进展》（ACS）会议（2022年）。[9] 维基媒体基金会, 《自助结账》， [访问时间：1月。 04, 2023[引用日期：04.01。2022]网址：https://en.wikipedia.org/wiki/Self-checkout [10] 维基媒体基金会, 《多车碰撞》， [访问时间：1月。04, 2023] [引用日期：04.01。2022]网址：https://en.wikipedia.org/wiki/Multiple-vehicle_collision[11]B. Goertzel、C. Pennachin, 《通用人工智能》，第2卷，施普林格，2007年。[12] T. Senator, 《开放世界新颖性的人工智能与学习科学》（SAIL-ON）（2019年）[引用日期：10.11。2022]网址：https://www.darpa.mil/program/science-of-artificial-intelligence-and-learning-for-open-world-novelty[13] P. Langley, 《面向彻底自主智能体的开放世界学习》，载于：《AAAI 人工智能会议论文集》，第34卷，2020年，第13539–13543页。

[14] M. Kejriwal、S. Thomas，用于在大富翁中生成新颖性的多智能体模拟器，《仿真建模实践与理论》112（2021）102364。doi:https://doi.org/10.1016/j.simpat.2021.102364。URL https://www.sciencedirect.com/science/article/pii/S1569190X21000770[15] P. Feeney、S. Schneider、P. Lymperopoulos、L. Liu、M. Scheutz、M. C. Hughes，NovelCraft：用于开放世界中新颖性检测与发现的数据集，arXiv预印本 arXiv:2206.11736（2022）。[16] CartPole，CartPole三维领域（2022）。URL https://github.com/holderlb/WSU-SAILON-NG/tree/master/domains/cartpole[17] J. Renz、X. Ge、S. Gould、P. Zhang，愤怒的小鸟人工智能竞赛，《人工智能杂志》36（2015）85–87。doi:10.1609/aimag.v36i2.2588。[18] J. Renz、X. Ge、M. Stephenson、P. Zhang，人工智能遇上愤怒的小鸟，《自然·机器智能》1（7）（2019）328–328。[19] C. Xue、V. Pinto、P. Zhang、C. Gamage、E. Nikonova、J. Renz，Science Birds Novelty：物理领域的开放世界学习测试平台，AAAI人工智能会议论文集，《开放世界人工智能设计》（2022）。[20] C. Gamage、V. Pinto、C. Xue、M. Stephenson、P. Zhang、J. Renz，愤怒的小鸟风格物理游戏的人工智能智能体新颖性生成框架，载于：2021 IEEE游戏会议，COG 2021，2021。[21] V. Pinto、C. Xue、C. N. Gamage、J. Renz，开放世界物理领域中的新颖性检测难题：以愤怒的小鸟为例，arXiv预印本 arXiv:2106.08670（2021）。[22] R. Riochet、M. Y. Castro、M. Bernard、A. Lerer、R. Fergus、V. Izard、E. Dupoux，Intphys 2019：视觉直觉物理理解的基准，ArXiv abs/1803.07616（2020）。[23] K. Yi*、C. Gan*、Y. Li、P. Kohli、J. Wu、A. Torralba、J. B. Tenenbaum，Clevrer：用于视频表征与推理的碰撞事件，载于：国际学习表征会议，2020。URL https://openreview.net/forum?id=HkxYzANYDB

[24] F. B. N. N. J. M. G. M. C. Wolf，COPHY：物理动力学的反事实学习，载于：ICLR，2020。doi:10.21227/ps5q-8m55。URL https://dx.doi.org/10.21227/ps5q-8m55[25] A. Bakhtin、L. van der Maaten、J. Johnson、L. Gustafson、R. Girshick，PHYRE：物理推理的新基准，载于：NeurIPS，2019。[26] D. Silver、J. Schrittwieser、K. Simonyan、I. Antonoglou、A. Huang、A. Guez、T. Hubert、L. Baker、M. Lai、A. Bolton 等，无需人类知识掌握围棋，《自然》550（7676）（2017）354–359。[27] D. Silver、T. Hubert、J. Schrittwieser、I. Antonoglou、M. Lai、A. Guez、M. Lanctot、L. Sifre、D. Kumaran、T. Graepel 等，通过自我对弈掌握国际象棋、将棋和围棋的通用强化学习算法，《科学》362（6419）（2018）1140–1144。[28] O. Vinyals、I. Babuschkin、W. M. Czarnecki、M. Mathieu、A. Dudzik、J. Chung、D. H. Choi、R. Powell、T. Ewalds、P. Georgiev 等，利用多智能体强化学习达到星际争霸II的大师级水平，《自然》575（7782）（2019）350–354。[29] SAIL-ON-BBA，广泛机构公告：开放世界新颖性的人工智能与学习科学（SAIL-ON），[访问日期：1月17, 2023]（2019年）。URL https://sam.gov/opp/88fdca99de93ddbb74cd8fb51916ceaa/ view[30] T. Boult、P. A. Grabowicz、D. Prijatelj、R. Stern、L. Holder、J. Alspector、M. Jafarzadeh、T. Ahmad、A. R. Dhamija、Cli、S. Cruz、A. Shrivastava、C. Vondrick、W. Scheirer，迈向新颖性形式理论的统一框架，AAAI人工智能会议论文集35（2021）15047–15052。[31] K. Doctor、C. Task、E. Kildebeck、M. Kejriwal、L. Holder、R. Leong，迈向跨领域定义领域复杂度度量，AAAI人工智能会议论文集，《开放世界人工智能设计》（2022）。[32] M. Molineaux、D. Dannenhauer，基于环境转换的开放世界学习智能体比较框架，AAAI人工智能会议论文集，《开放世界人工智能设计》（2022）。

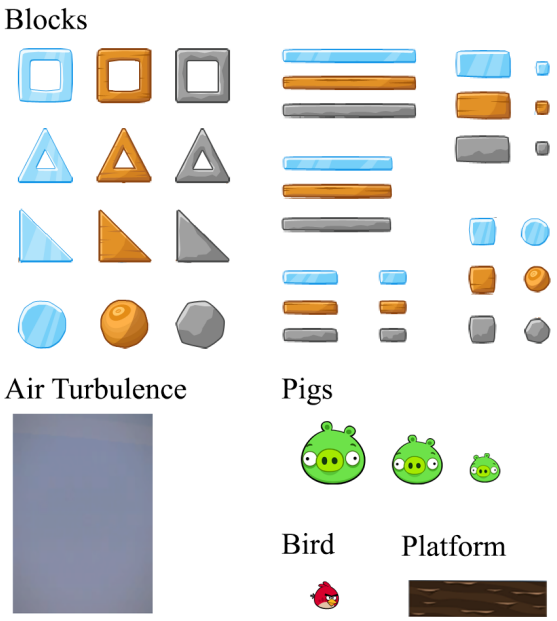
[33] J. Balloch、Z. Lin、M. Hussain、A. Srinivas、R. Wright、X. Peng、J. Kim、M. Riedl，《Novgrid：用于评估智能体对新颖性响应的灵活网格世界》，arXiv 预印本 arXiv:2203.12117（2022）。[34] S. Goel、G. Tatiya、M. Scheutz、J. Sinapov，《NovelGridworlds：用于检测和适应开放世界中新颖性的基准环境》，自主智能体与多智能体系统国际基金会，AAMAS（2021）。[35] M. Chevalier-Boisvert、L. Willems、S. Pal，《用于 Gymnasium 的极简网格世界环境》（2018）。URL https://github.com/Farama-Foundation/Minigrid[36] Minecraft，Minecraft 官方游戏（2022）。URL https://www.minecraft.net/en-us[37] R. A. Smaldone、C. M. Thompson、M. Evans、W. Voit，《通过电子游戏教授科学》，《自然·化学》9（2）（2017）97–102。[38] Rovio Enternainment，《愤怒的小鸟》游戏，[访问日期：12 月 07 日，2022] [引用日期 07.12。2022]。URL https://www.rovio.com/games/angry-birds[39] V. Pinto、J. Renz、C. Xue、P. Zhang、K. Doctor、D. W. Aha，《测量开放世界人工智能系统的性能》，AAAI人工智能会议论文集，《开放世界人工智能设计》（2022）。[40] M. Jafarzadeh、A. R. Dhamija、S. Cruz、C. Li、T. Ahmad、T. E. Boult，《开放世界学习综述及迈向无标签开放世界学习的步骤》，arXiv 电子论文（2020）arXiv–2011。[41] X. Peng、J. C. Balloch、M. O. Riedl，《游戏中的新颖性检测与适应》，arXiv 预印本 arXiv:2106.02204（2021）。[42] F. Muhammad、V. Sarathy、G. Tatiya、S. Goel、S. Gyawali、M. Guaman、J. Sinapov、M. Scheutz，《面向变化世界的以新颖性为中心的智能体架构》，载于：第20届国际自主智能体与多智能体系统会议论文集，2021年，第925–933页。[43] L. Ferreira、C. Toledo，《生成〈愤怒的小鸟〉关卡的基于搜索的方法》，载于：2014 IEEE计算智能与游戏会议，2014年，第1–8页。

[44] A. Sanborn、V. Mansinghka、T. Griffiths，《为碰撞物体调和直觉物理学与牛顿力学》，《心理学评论》120（2013年3月）。doi:10.1037/a0031912。[45] J. Bliss、J. Ogborn，《从起点学习力与运动》，《学习与教学》4（1）（1994）7–25。doi:https://doi.org/10.1016/0959-4752(94)90016-7。URL h
ttps://www.sciencedirect.co
m/science/article/pii/0959475294900167[46] AIBIRDS，《愤怒的小鸟》人工智能竞赛 [引用日期 22.12。2022]。URL http://aibirds.org/[47]T. Borovi˘cka、R. ˘Spetl´ık、K. Ryme˘s，Datalab 愤怒的小鸟人工智能 [引用日期 22.12。2022]。URL http://aibirds.org/2014-papers/datalab-birds.pdf[48] T. J. Wang，AI Angry Birds Eagle Wing [引用日期 22.12。2022]。URL https://github.com/heartyguy/AI-AngryBird-Eagle-Wing[49] M. Stephenson、J. Renz、X. Ge、P. Zhang，《2017年 AIBIRDS 竞赛》，ArXiv abs/1803.05156（2018）。[50] Z. Wang、T. Schaul、M. Hessel、H. Hasselt、M. Lanctot、N. Freitas，《用于深度强化学习的对决网络架构》，载于：国际机器学习会议，PMLR，2016年，第1995–2003页。[51] H. Van Hasselt、A. Guez、D. Silver，《采用双重 Q 学习的深度强化学习》，载于：AAAI人工智能会议论文集，第30卷，2016年。[52] V. Zambaldi、D. Raposo、A. Santoro、V. Bapst、Y. Li、I. Babuschkin、K. Tuyls、D. Reichert、T. Lillicrap、E. Lockhart 等，《关系型深度强化学习》，arXiv 预印本 arXiv:1806.01830（2018）。[53] C. Xue、E. Nikonova、P. Zhang、J. Renz，《通过适应原则学习实现快速开放世界适应》，提交至《人工智能期刊》（2023）。[54] O. Ahmed*、F. Tr¨auble*、A. Goyal、A. Neitz、Y. Bengio、B. Sch¨olkopf、M. W¨uthrich、S. Bauer，《CausalWorld：用于因果结构与迁移学习的机器人操作基准》，载于：第9届国际

学习表征会议（ICLR），2021年，*表示共同贡献。网址
https://openreview.net/pdf?id=SK7A5pdrgov

[55] S. James、Z. Ma、D. R. Arrojo、A. J. Davison，Rlbench：机器人学习基准与学习环境，《IEEE机器人与自动化快报》5（2）（2020年）3019–3026。

附录A. NovPhy 任务中的物体



图A.10：NovPhy 常规任务中出现的物体（未按比例绘制）。除平台物体可呈现不同形状和尺寸外，其他物体的形状和尺寸均固定。

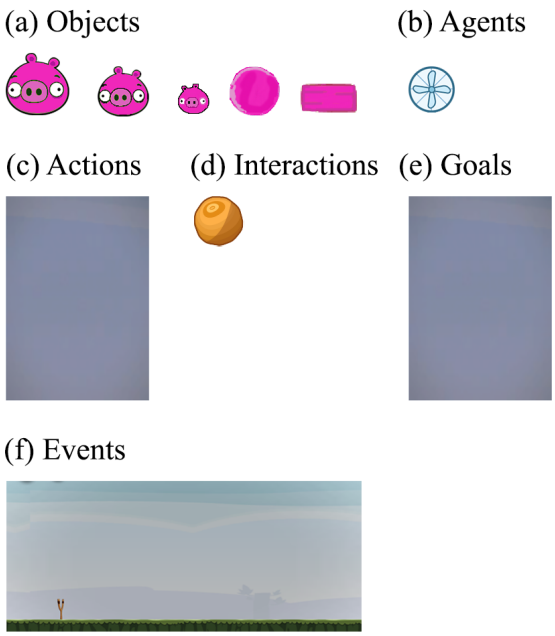


图 A.11：NovPhy 中与新颖性相关的物体（未按比例绘制）。此处未展示关系（通常位于任务左侧的弹弓现在位于任务右侧）和环境（环境中的重力现在反向）新颖性，因为它们并不与特定的游戏物体相关。图中展示了：(a) 物体：一种不同于普通猪/方块颜色（粉色）的新猪/方块；(b) 智能体：一种新颖的外部智能体——风扇，可吹动空气（从左向右水平吹动），从而影响物体的运动路径；(c) 动作：空气湍流外部智能体增大其向上作用力的大小（外观与空气湍流智能体相同）；(d) 交互：现有的圆形木质物体现在具有磁性：排斥同类型物体，并吸引其他类型的物体（外观与圆形木质物体相同）；(e) 目标：空气湍流外部智能体将其目标从向上推动物体改为向下推动物体（外观与空气湍流智能体相同）；(f) 事件：当第一只小鸟死亡时，会发生一场风暴，对右方施加力，从而影响物体的运动（图中展示了风暴发生时环境的外观）。

附录 B. NovPhy 中的任务

NovPhy 包含 40 个新颖任务模板和相应的 40 个常规任务模板，这些模板通过应用八种新颖性，为五种物理场景设计。图 B.12、B.13、B.14、B.15 和 B.16 分别展示了五种场景（单一力、多个力、滚动、下落和滑动）的常规任务模板和新颖任务模板。图 B.17 展示了根据五个示例任务模板（物体新颖性）生成的任务。

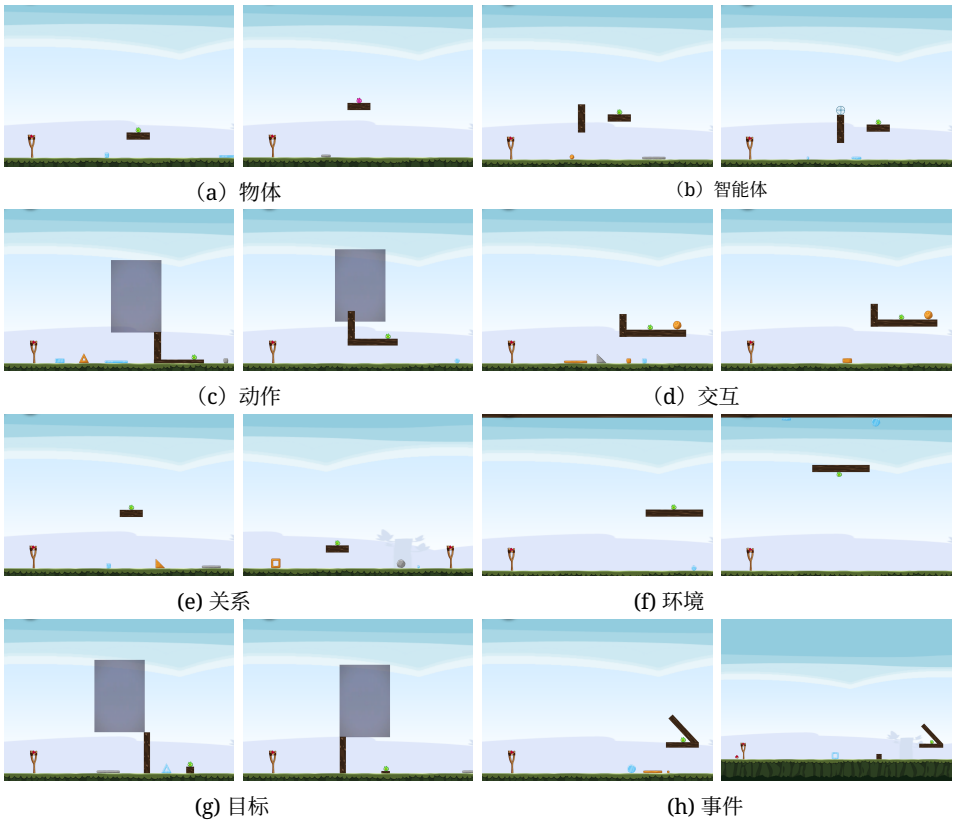
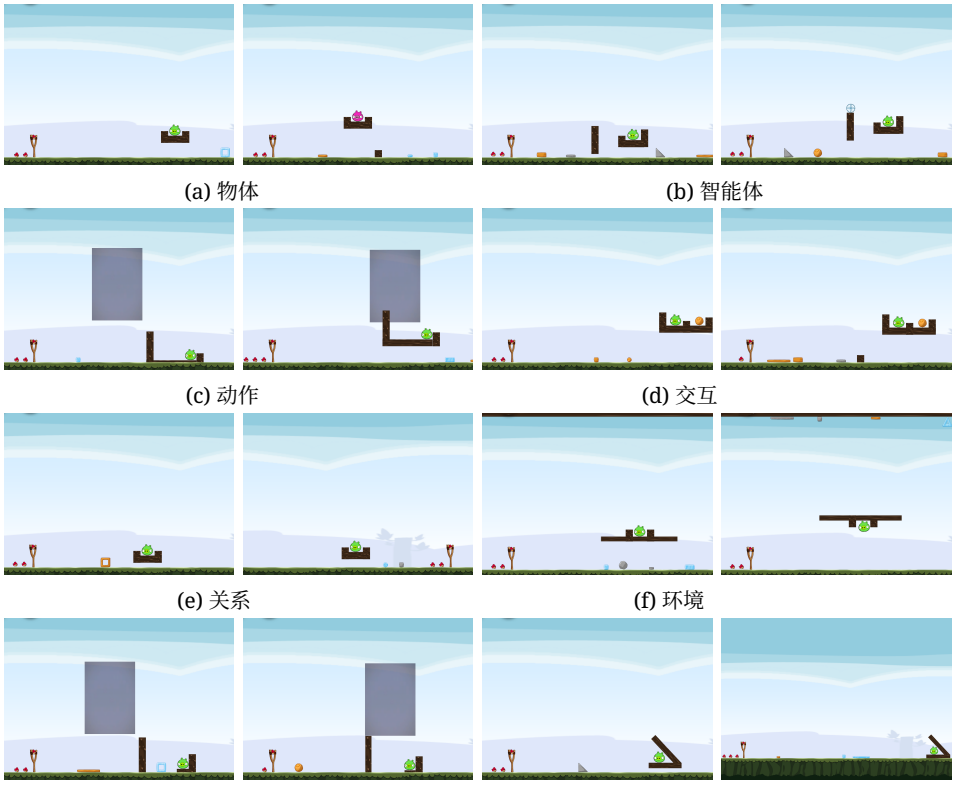


图 B.12：应用了八种新颖性的单力场景任务模板。在每个子图中，左侧图示为常规任务，右侧图示为应用了相应新颖性的对应新颖任务。



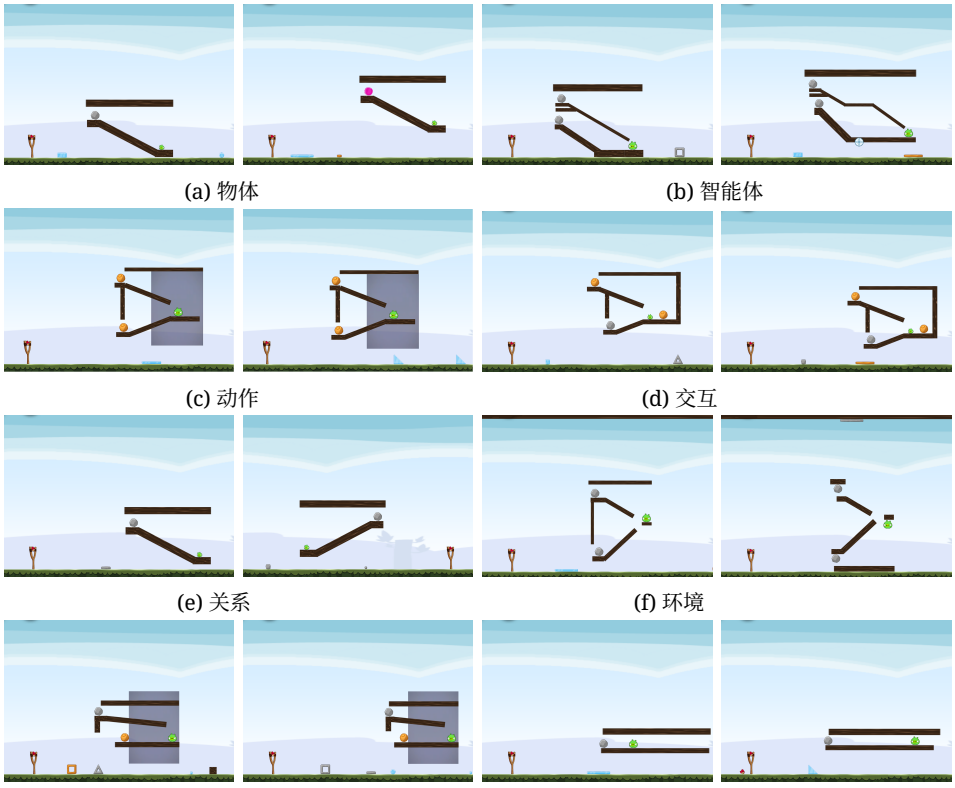
(a) 物体　(b) 智能体

(c) 动作　(d) 交互

(e) 关系　(f) 环境

(g) 目标　(h) 事件

图 B.13：应用了八种新颖性的多种力场景任务模板。在每个子图中，左侧图展示常规任务，右侧图展示应用了新颖性的对应新颖任务。



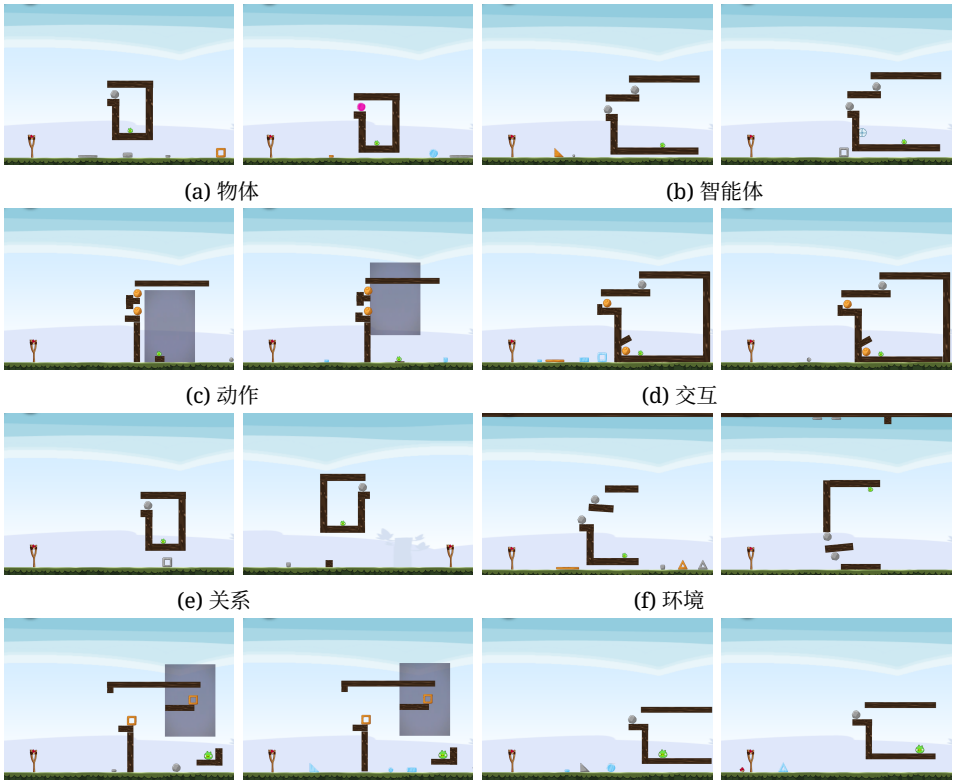
(a) 物体　(b) 智能体

(c) 动作　(d) 交互

(e) 关系　(f) 环境

(g) 目标　(h) 事件

图 B.14：应用了八种新颖性的滚动场景任务模板。每个子图中，左侧图为常规任务，右侧图为应用了新颖性的对应新颖任务。



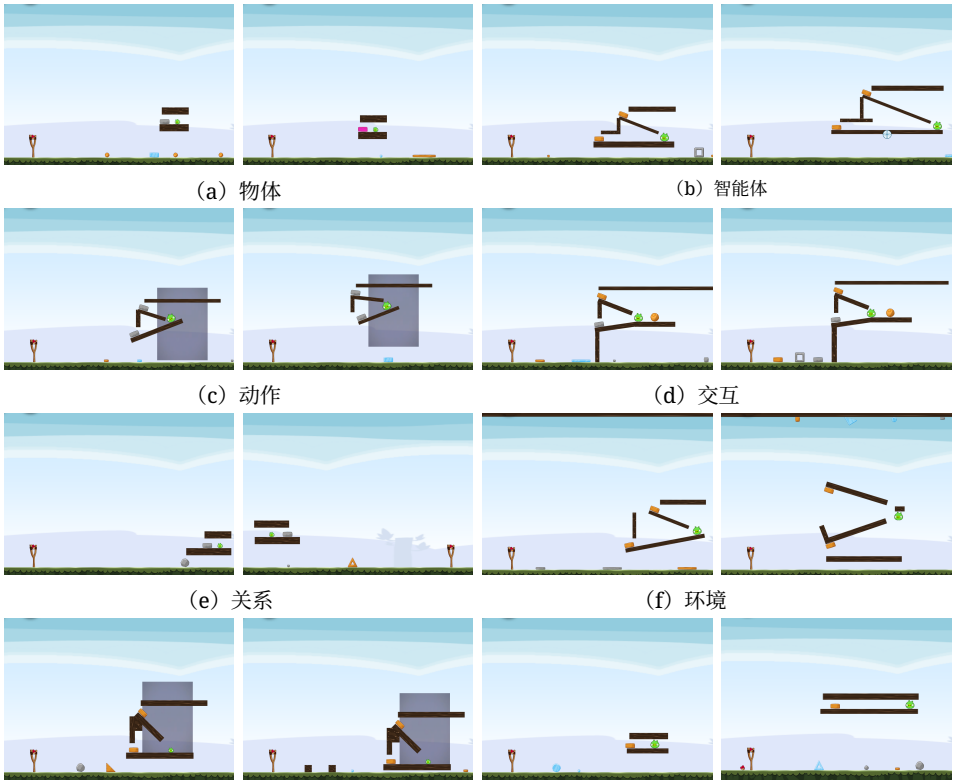
(a) 物体 (b) 智能体

(c) 动作 (d) 交互

(e) 关系 (f) 环境

(g) 目标 (h) 事件

图 B.15：应用了八种新颖性的下落场景任务模板。每个子图中，左图为常规任务，右图为应用相应新颖性后的对应新颖任务。



（a）物体　（b）智能体

（c）动作　（d）交互

（e）关系　（f）环境

（g）目标　（h）事件

图 B.16：应用了八种新颖性的滑动场景任务模板。在每个子图中，左图为常规任务，右图为应用了相应新颖性的对应新颖任务。

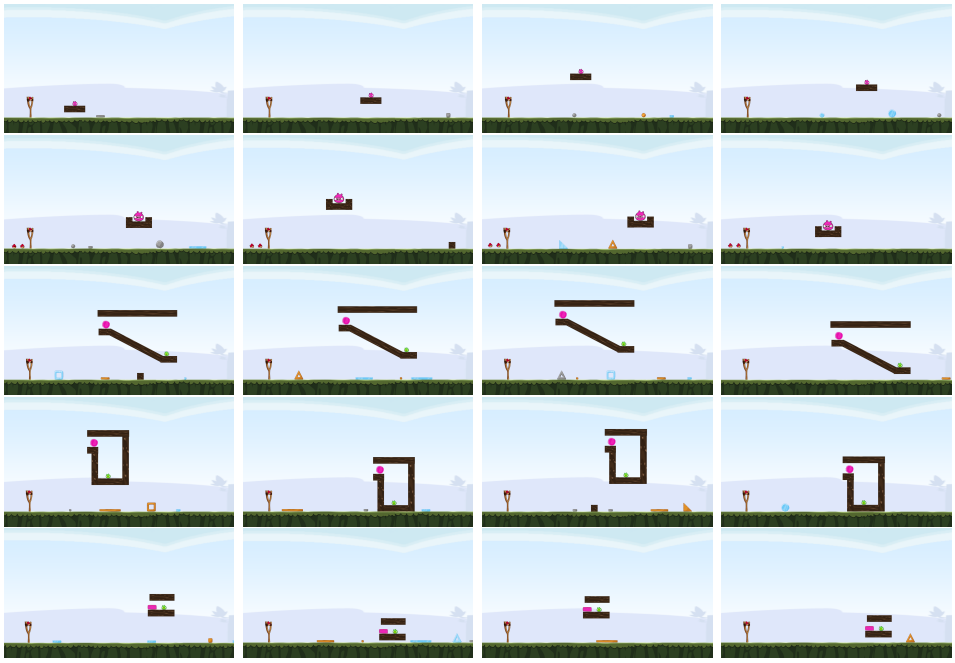


图 B.17：每行展示了由同一任务模板生成的四个任务示例。同一模板下各任务中的物体和干扰对象的位置各不相同，因此每个任务都有各自的解；然而，同一模板下的所有任务都可以通过该任务模板所属物理场景中的同一物理规则来解决。

附录 C。NovPhy 中新颖任务的设计

正如主论文第 3.1 节所述，在设计新颖任务时，我们遵循以下要求：智能体必须在新颖性所产生的影响下完成任务。为满足这一要求，在设计新颖任务时，我们确保将新颖性引入任务，使其至少影响任务解中的一个物理交互阶段（初始阶段、中间阶段或最终阶段）。表 C.4 展示了在 NovPhy 中设计 40 个新颖性场景时受到影响的物理交互阶段。

物理场景	新颖性	受影响的物理交互阶段
单一力	1	最终
	2	中间
	3	中间
	4	中间
	5	初始，中间
	6	初始，中间
	7	中间
	8	中间
多个力	1	最终
	2	中间
	3	中间 ，最终
	4	中间 ，最终
	5	初始，中间
	6	初始，中间
	7	中间 ，最终
	8	中间
滚动	1	初始
	2	中间
	3	中间
	4	中间
	5	初始，中间
	6	初始，中间
	7	中间
	8	中间
下落	1	初始
	2	中间
	3	中间
	4	中间
	5	初始，中间
	6	初始，中间
	7	中间
	8	中间
滑动	1	初始
	2	中间
	3	中间
	4	中间
	5	初始，中间
	6	初始，中间
	7	中间
	8	中间

表 C.4：表格
受新颖性影响的解的物理交互阶段。新颖性 1 至 8 分别代表：1. 物体，2. 智能体，3. 动作，4. 交互，5. 关系，6. 环境，7. 目标，以及 8. 事件。

附录 D. 新颖性检测性能

检测性能仅在无信息条件下运行的智能体中进行测量：Datalab、Eagle's Wing、Pig Shooter 和 Random。智能体的检测模块基于通过率偏差。我们纳入了两种算法：1）简单移动平均通过率（sma）；2）预设移动平均通过率（pma），这两种算法均可用作检测模块。

简单移动平均方法比较给定窗口大小下两个连续的移动平均通过率，并检查第二个通过率是否相对于第一个通过率偏离了超过所选阈值。我们研究了窗口大小 5 和 10，以及阈值 0.2、0.4、0.6 和 0.8 的变化情况。

在预设移动平均方法中，我们假设经过训练的智能体知道给定任务模板的常规任务通过率，因为这些任务事先已提供给智能体进行训练。因此，在该方法中，我们检查给定窗口大小内的通过率是否偏离了若干个标准差（检测阈值）。我们研究了窗口大小 5 和 10，以及阈值 1.5 和 2 的变化情况。



图 D.18：两种算法 sma 和 pma 下智能体的 CDT，其中窗口大小发生变化（算法名称后的第一个值），检测阈值也发生变化（最后一个值）。

图 D.18 展示了智能体在不同窗口大小和检测阈值下的 CDT 结果。对于 Datalab、Eagle's Wing 和 Pig Shooter，窗口大小为 10、检测阈值为 1.5 个标准差的 pma 方法取得了最高的 CDT；对于随机智能体，窗口大小为 5、检测阈值为 0.2 的 sma 方法取得了最高的 CDT。取得最高 CDT 的方法被用作相应智能体的新颖性检测方法，而该方法取得的结果

已在主论文中呈现，并用于附录其余部分。

新颖性检测性能使用 CDT 和 DD 指标进行衡量。图 D.19-D.22 展示了各智能体的 CDT（越高越好）和 DD（越低越好）热力图。图 D.23 和 D.24 展示了按新颖性汇总的检测性能（NPS_{CDT}），图 D.25 和 D.26 则展示了按场景汇总的结果（SPN_{CDT}）。从这四幅汇总图以及智能体热力图与人类表现的对比中可以看出，智能体在新颖性检测方面仍有很大的提升空间。

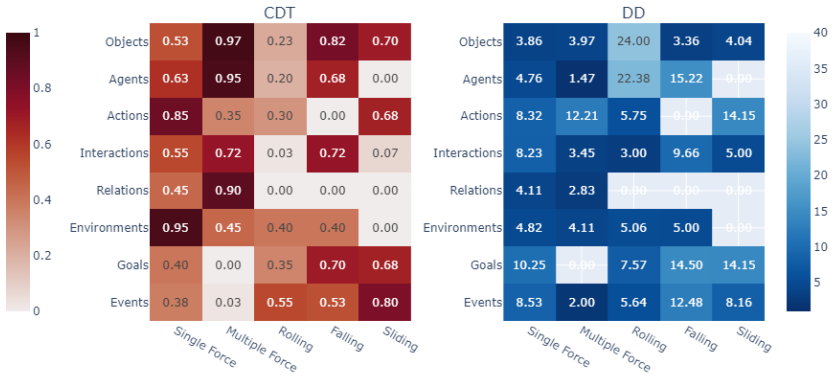


图 D.19：Datalab 的 CDT 和 DD 结果。

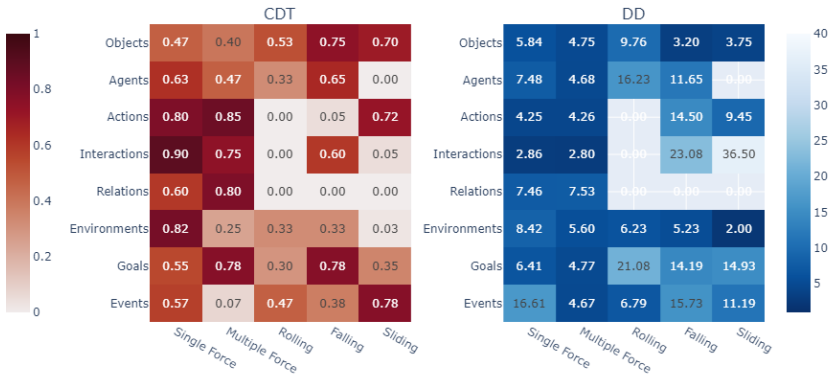


图 D.20：Eagle's Wing 的 CDT 和 DD 结果。

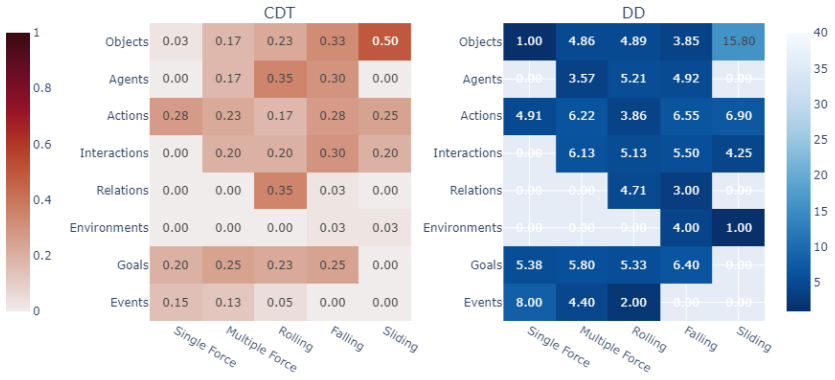


图 D.21：Pig Shooter 的 CDT 和 DD 结果。

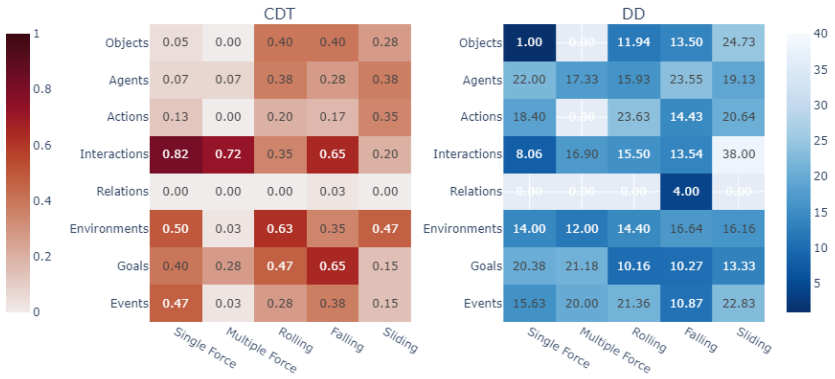


图 D.22：随机智能体的 CDT 和 DD 结果。

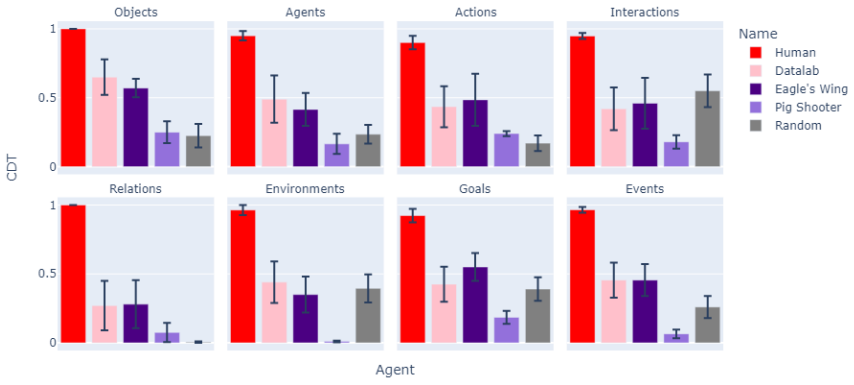


图 D.23： NPS_{CDT} 八种新颖性对应的智能体。

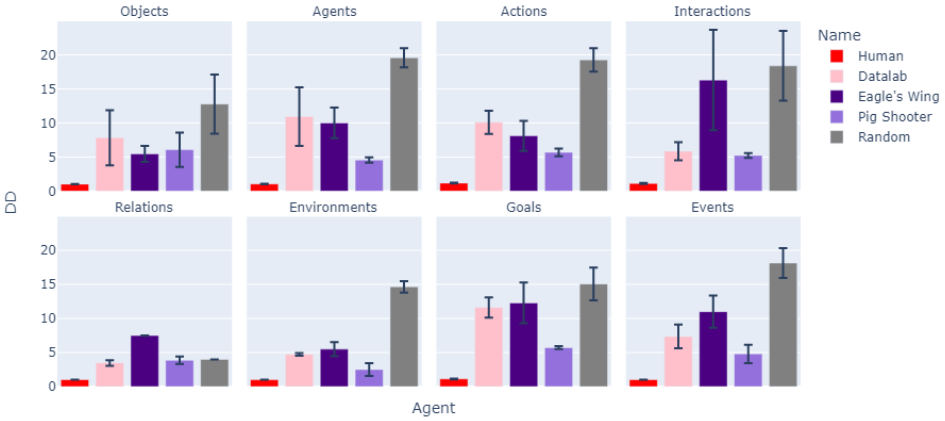


图 D.24： NPS_{DD} 智能体在八种新颖性方面的表现。

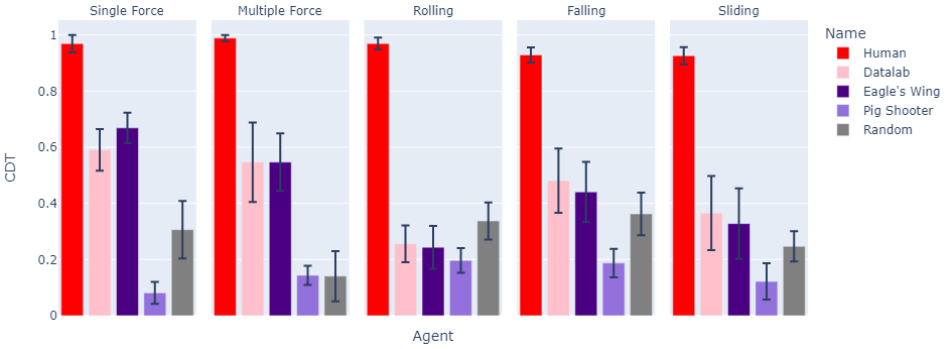


图 D.25： SPN_{CDT} 智能体在五种场景中的表现。

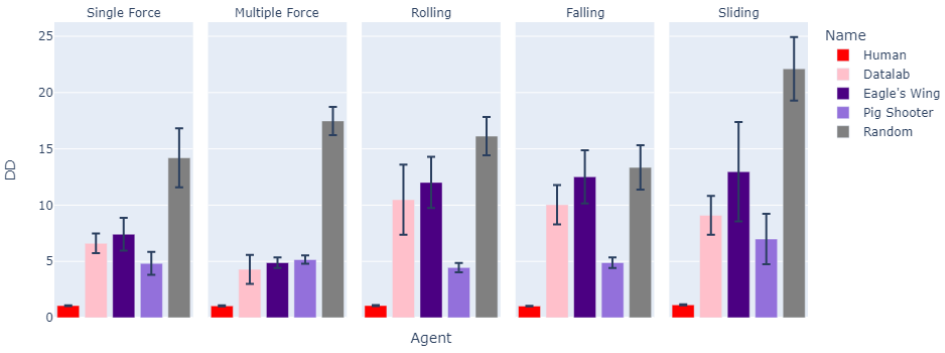


图 D.26： SPN_{DD} 智能体在五种场景中的表现。

附录 E. 新颖性适应性能



图 E.27：智能体在所有新颖性场景中的通过率。横轴表示试验中任务的索引。索引 -40 至 -1 对应常规任务，0 至 40 对应新颖任务。纵轴表示通过率。

各新颖性场景的适应曲线如图 E.27 所示。根据各智能体的适应曲线得出的 AP 和 AUS 值，展示在图 E.28 至 E.38 的热力图中。AUS 值根据智能体在新颖任务最后 50%（m=20）中的性能计算得出。各新颖性对应的适应结果见图 E.39 和 E.40，而图 E.41 和 E.42 展示了各场景性能对应的适应结果。可以看出，智能体 DQN Adapt 和 DQN Relational 在部分新颖性场景中能够达到人类 AP 和 AUS 结果。然而，从适应图来看，在与人类接受相同数量任务的情况下，这两个智能体尚未

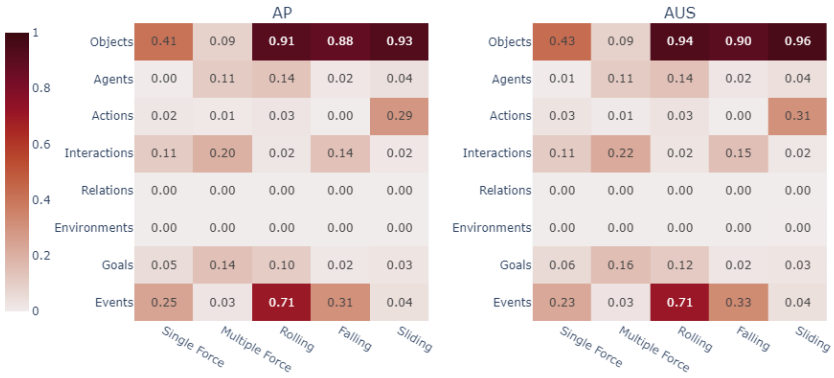


图 E.28：DQN 离线的平均精度和 AUS 结果。

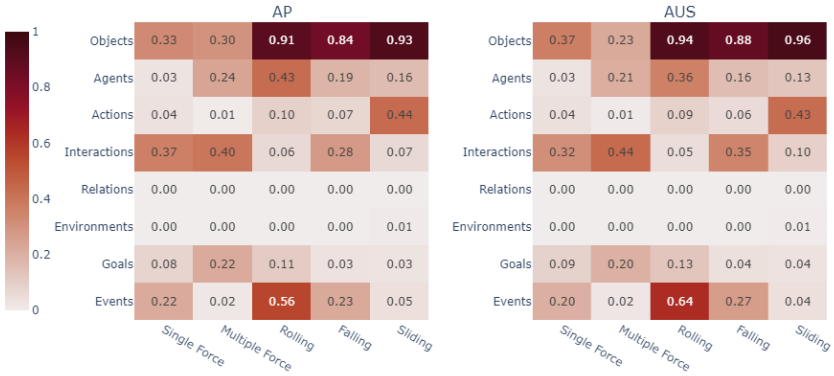


图 E.29：DQN 在线的平均精度和 AUS 结果。

达到人类能够实现的性能，这表明其适应效率仍有提升空间。

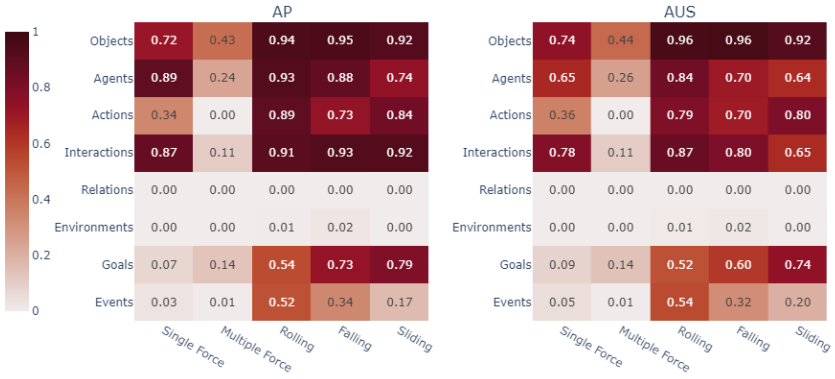


图 E.30：DQN Adapt 的平均精度和 AUS 结果。

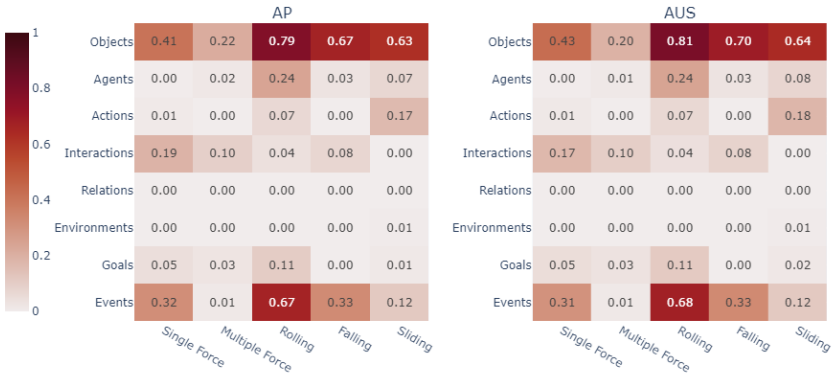


图 E.31：关系离线的平均精度和 AUS 结果。

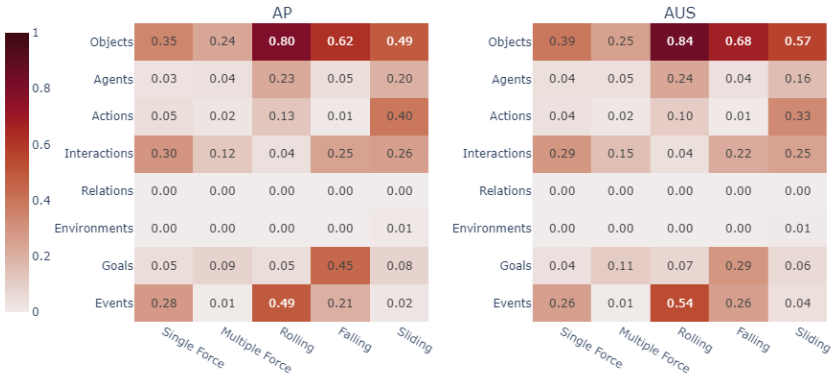


图 E.32：关系在线的平均精度和 AUS 结果。

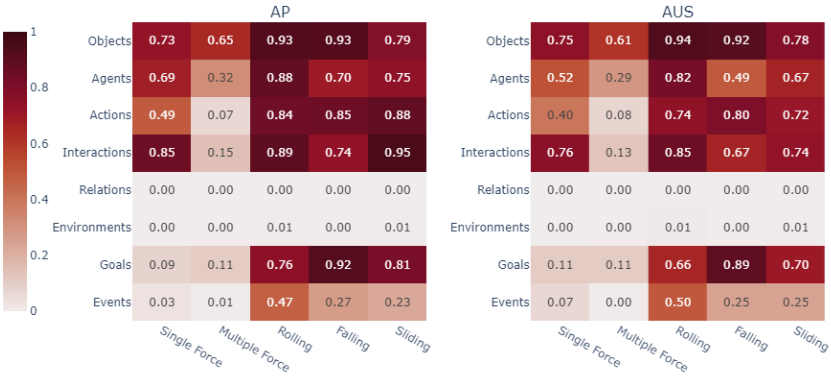


图 E.33：Relational Adapt 的平均精度和 AUS 结果。

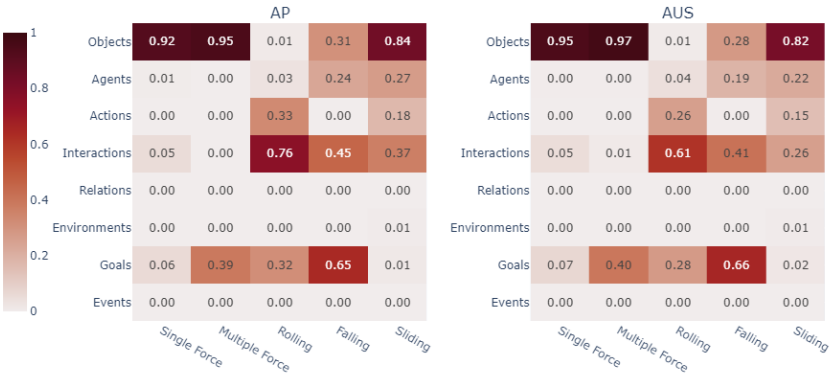


图 E.34：朴素自适应的平均精度和 AUS 结果。

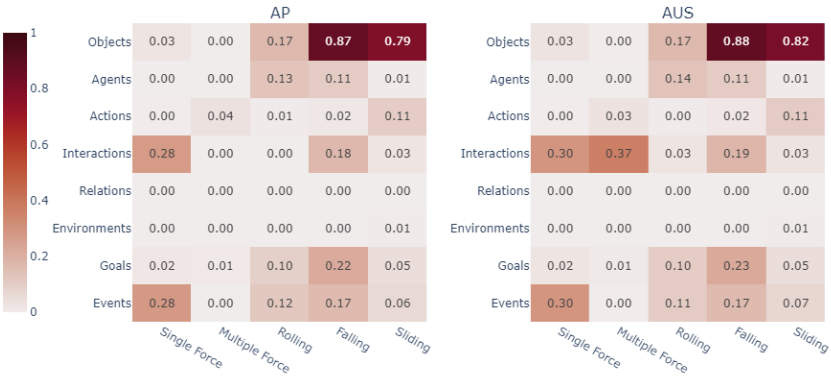


图 E.35：Datalab 的平均精度和 AUS 结果。

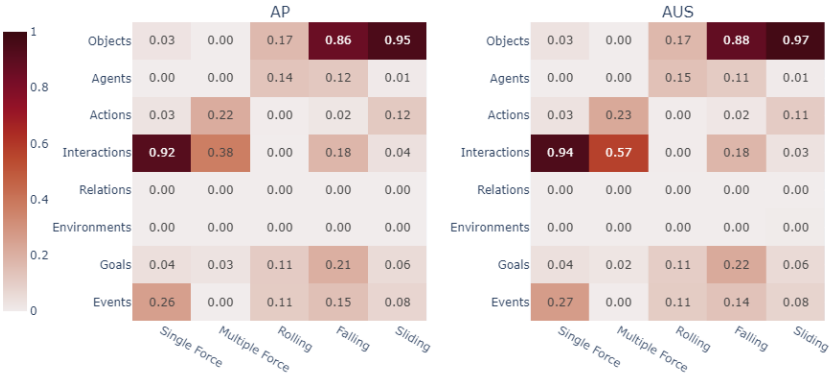


图 E.36：Eagle's Wing 的平均精度和 AUS 结果。

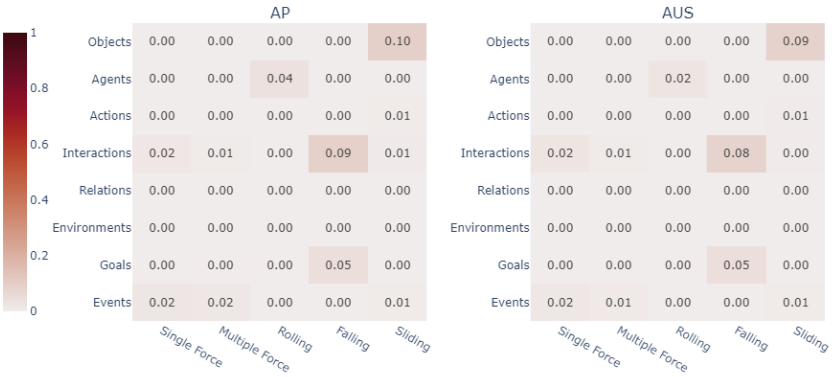


图 E.37：Pig shooter 的平均精度和 AUS 结果。

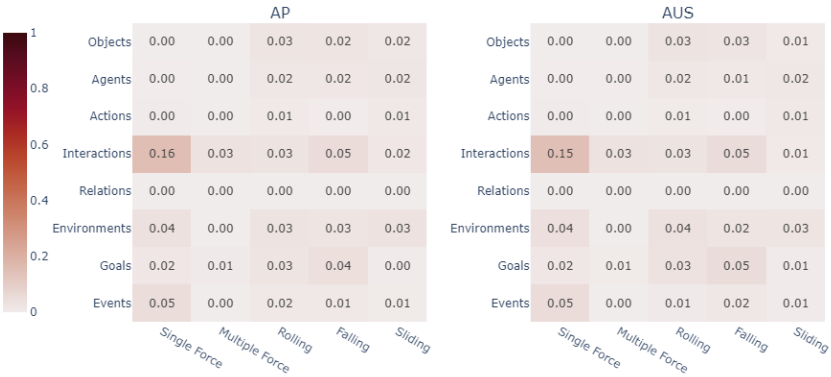


图 E.38：随机智能体的平均精度和 AUS 结果。

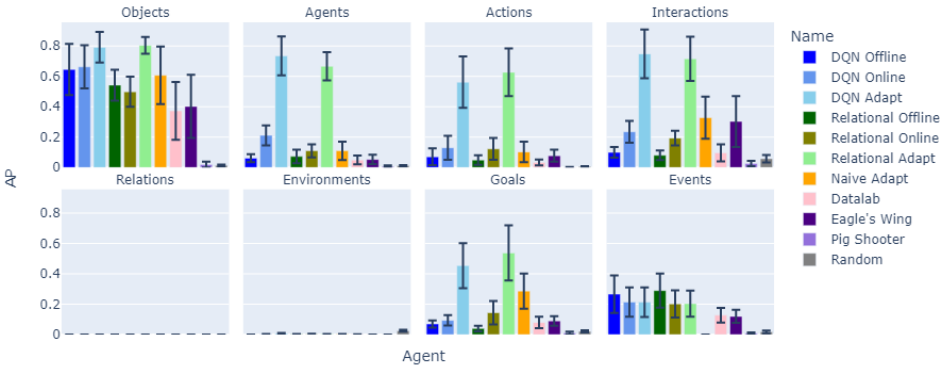


图 E.39：八种新颖性下智能体的 NPS_{AP}。

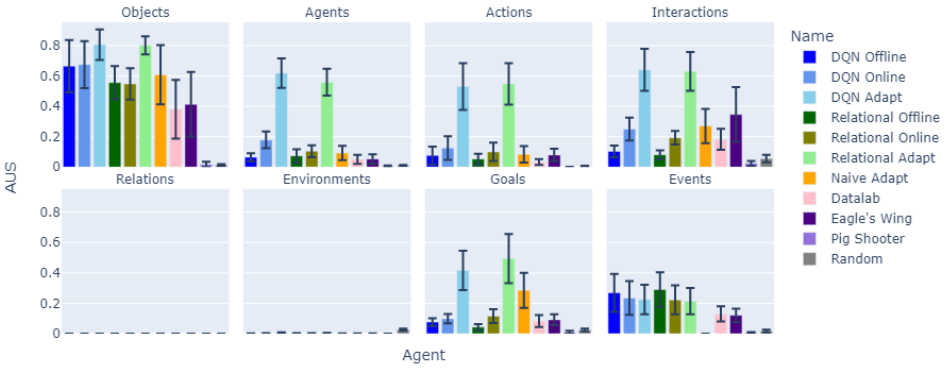


图 E.40：八种新颖性下智能体的 NPS_{AUS}。

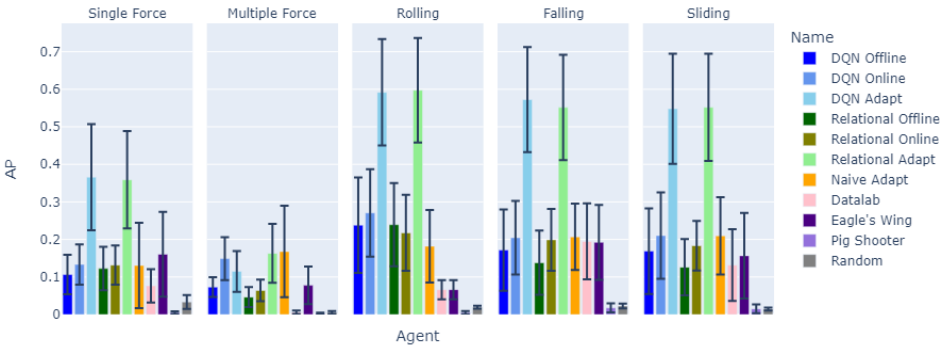


图 E.41：五种场景下智能体的 SPN_{AP}。

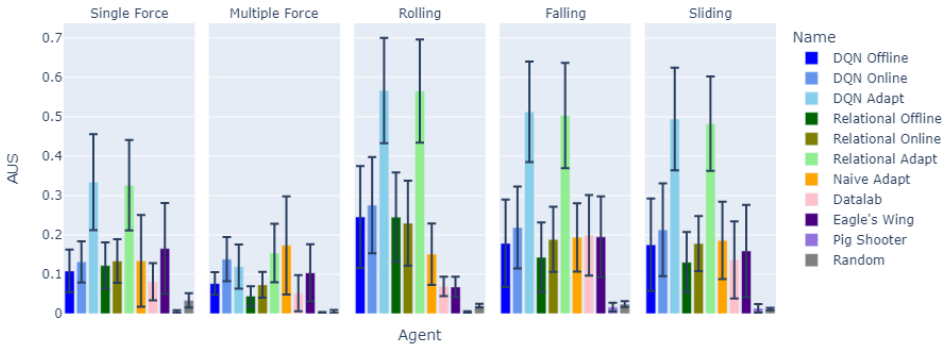


图 E.42：五个场景中智能体的 SPN_{AUS} 。

附录 F. 人类检测与适应

在前几节中，我们讨论了人类的检测与适应性能，并表明人类的性能远优于智能体的性能。本节分析人类的新颖性检测与适应性能之间的关系。

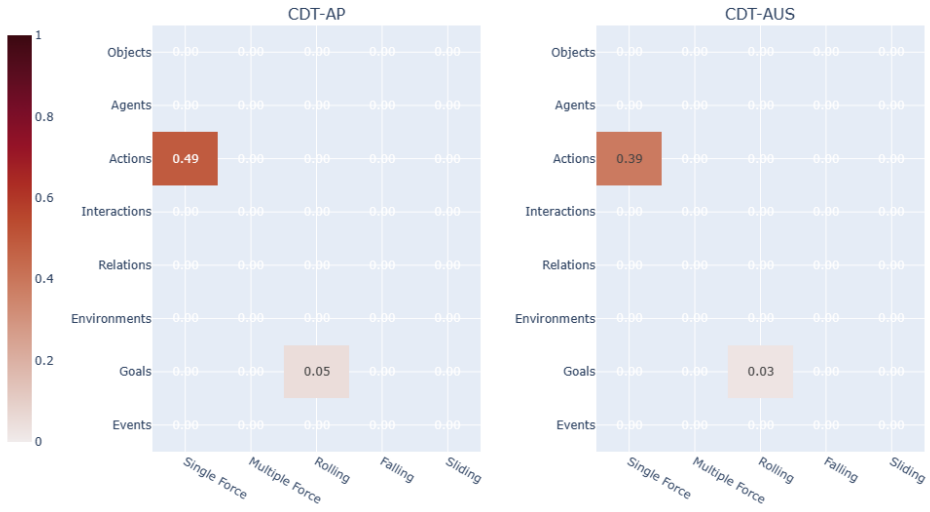


图 F.43：CDT=1 组与 CDT=0组适应性能的曼-惠特尼检验 p 值

图 F.43 展示了曼-惠特尼检验的 p 值，该检验用于测试个人是否正确检测到新颖性（CDT $=1$ 或 CDT $=0$）是否会导致适应性能存在差异。对于

CDT $= 0$ 组，我们仅使用了未检测到新颖性的玩家数据（排除了误报玩家的数据）。零假设是两组的适应性能相同。如图 F.43 所示，几乎所有条目都未定义，因为所有人类玩家都正确检测到了新颖性，因此在这些新颖性场景中没有未检测到新颖性的玩家。对于动作-单一力新颖性场景，平均精度和 AUS 的 p 值均超过了我们设定的 5% 显著性水平。也就是说，我们没有足够的证据拒绝零假设，这意味着无论是否检测到新颖性，玩家的表现都相似。这里的新颖性是空气湍流向上作用力的增加，除非进行交互，否则无法通过视觉察觉。一些人类玩家没有检测到这种新颖性，但仍成功完成了任务，因为从常规任务到新颖任务，动作只有细微变化。另一方面，在目标-滚动新颖性场景中，适应性能会因 CDT 而存在显著差异。在这一新颖性场景中，从常规任务转为新颖任务时，必须大幅调整动作，因此如果没有检测到新颖性，就很难适应。

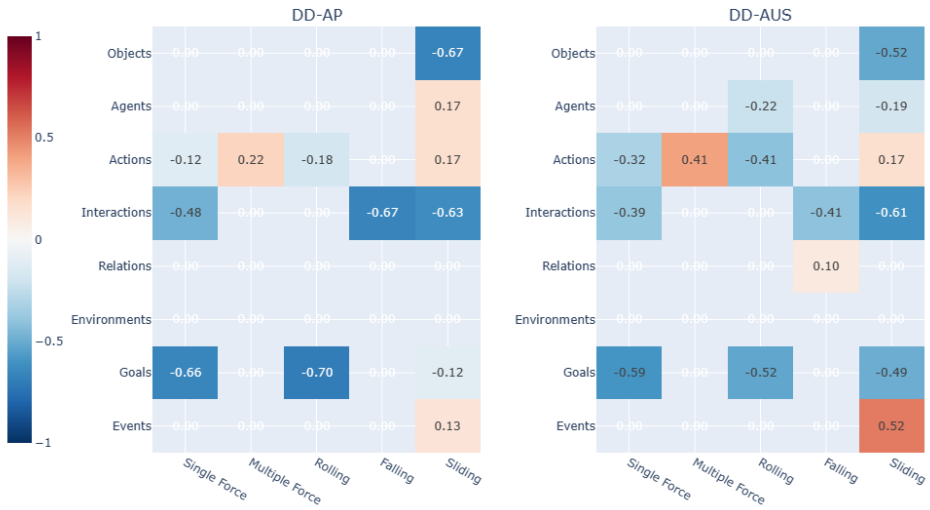


图 F.44：适应性能与检测延迟的斯皮尔曼秩相关

图 F.44 展示了检测延迟与适应性能（平均精度和 AUS）之间的斯皮尔曼秩相关系数。在此次计算中，如果参与者在试验中未检测到新颖性，我们假设检测延迟为五，即一次试验中新颖实例的最大数量

。如图所示，对于某些新颖性场景，相关系数未定义。这些情况是所有人类玩家都在第一个新颖任务中检测到新颖性的情形（DD $= 1$）。如 DD-AP 热力图所示，存在五个中等程度的负相关（介于 -0.6 和 -0.7 之间），且在 5% 的显著性水平下均具有统计显著性。这表明，对于这些新颖性场景，检测所需的任务越多，渐近适应性能越低。根据 DD-AUS 热力图，两个新颖性场景具有中等程度的正相关（0.41 和 0.52）。然而，它们在 5% 的显著性水平下均不具有统计显著性。只有新颖性场景交互-滑动中的中等程度强负相关值在 5% 的显著性水平下具有统计显著性。