arXiv:2601.00324v1 [cs.AI] 1 Jan 2026

# 流动性博弈的多智能体强化学习

**Alicia Vidler**[1],**Gal A. Kaminka**[1]

[1]巴伊兰大学 aliciavidler@gmail.com, galk@cs.biu.ac.il

**摘要**

将群体方法用于流动性金融市场建模，并将金融分析技术用于群体分析，有望推动这两个研究领域的发展。在群体研究中，运用博弈论方法有望解释集体效用遵循这一现象，并说明理性、自利的群体参与者如何做到这一点。在金融市场中，若能更好地理解相互独立的金融智能体如何自我组织，从而改善市场运行并增强其稳定性，将有助于市场设计研究者。本文将流动性博弈与理性群体结合起来：在流动性博弈中，交易者的收益取决于一次交易中的总流动性；在理性群体中，去中心化智能体使用差分奖励，使自利学习与全局目标保持一致。我们提出一个理论框架，定义了一群交易者，其集体目标是在保持智能体独立性的同时提供市场流动性。基于马尔可夫团队博弈框架运用差分奖励，我们证明，个体的流动性最大化行为有助于提升整体市场流动性，而无需协调或合谋。该金融群体模型为理性、独立的智能体提供了一个建模框架，使其在双边资产市场中既能实现个体盈利，又能提升集体市场效率。

## 引言

成交量以及系统高效实现这一成交量的能力，通常统称为流动性。在现代市场中，它是异质智能体之间去中心化交互所产生的涌现且理想的属性，而非固定不变的特征。流动性指的是能够快速、低成本地交易资产，且不会显著影响其价格的难易程度（Frank J. Fabozzi 2005）。借用美国联邦储备委员会理事会成员 Kevin Warsh 的话来说：“流动性就是信心。”

提高市场流动性是一项颇具挑战的问题。市场出清——将意图转化为已执行交易——源于无数局部选择（报价、库存、风险限额），这些选择会产生系统层面的影响（例如总交易量）。然而，由于在现实世界的现代金融市场中，智能体联盟与协调通常是非法的，因此这些局部选择必须彼此不协调——地进行。独立学习者（智能体）的交易意图（及其执行）必须依据其严格的局部奖励，而这些奖励与全市场目标之间并无明确联系。

本文探讨独立智能体自利的决策与市场整体交易量所期望实现的最大流动性之间的张力。具体而言，我们将流动性形成的博弈论模型流动性博弈（*LG*）（Vidler and Walsh 2024b）与一种去中心化多智能体强化学习方法理性群体（Kaminka 2025）相结合。在此框架中，多个金融智能体（投资者、银行或做市商）依据局部信息采取行动。它们的集体决策决定了系统层面的流动性模式。去中心化学习框架使独立学习者能够采取提升流动性的行动，同时保持各自的自利性。

具体而言，我们将市场系统建模为马尔可夫团队博弈，即完全合作的随机重复博弈。在这种博弈中，智能体理论上会因其联合行动而获得相同的全局（系统级）收益。我们将汇总清算成交量视为流动性的全局指标。这是一项系统级性能指标，用于衡量每个时段内实际执行（即匹配并结算）的订单数量，与报价量或提交量不同（Bank of England 2015）。汇总清算成交量是需要最大化的目标属性。根据（Kaminka 2025），我们应对这样一项挑战：智能体无法获得（观察到）这一联合收益，只能接触到它的局部代理指标。因此，每个智能体都通过局部差分奖励（对流动性的边际贡献）进行学习，从而获得与全局流动性相一致的个体奖励信号。这使得智能体能够完全独立地学习，无需协调编排或合谋。我们考察以下两个假设：

**[H1]：采用差分奖励的流动性**。在随机配对的双边交易中，使用差分奖励训练的智能体将比使用其他局部或全局信号的智能体，以及不进行学习的基线方法，获得更高的总流动性和个体交易量，从而无需协调即可最大化局部效用和全局效用。

**[H2]：精确匹配下的鲁棒性**。在精确匹配流动性博弈（无部分成交）中，使用差分奖励训练的智能体将学会实现并

版权所有 © 2026年，人工智能促进协会 (www.aaai.org)。保留所有权利。

保持较高的匹配效率，并且比采用其他奖励设计的智能体更早在训练过程中达到较高的交易成功阈值。

我们在不同的清算机制下检验这些假设。清算机制定义了交易报价如何转化为实际执行的交易，进而影响全局目标$G$（清算成交量，从而影响流动性）。我们探讨两种机制：（i）部分成交，允许按比例成交（我们称之为最小成交 机制）；（ii）精确匹配，要求买方和卖方的数量完全匹配（Vidler and Walsh 2024b）。

我们表明，采用差分奖励相较于其他学习方法和非学习方法具有优势。根据我们的结果，在部分清算规则（最小成交）下，差分奖励智能体实现了最高的总流动性和清算效率；在精确匹配下，尽管约束更严格，它们仍具有竞争力，并迅速达到 $\sim$ 70%的交易成功率。

本文结构如下。首先，我们探讨相关文献。然后，我们详细介绍方法和框架，将分布式多智能体强化学习中的差分奖励函数与侧重流动性的债券市场博弈论模型相结合。在接下来的章节中，我们将阐述这一组合范式的具体实例，并讨论研究结果。

## 金融市场作为多智能体系统：文献综述

传统经济理论采用了多种数学框架来模拟金融市场中的智能体交互，每种框架都从不同角度揭示了集体行为和市场动态。金融市场可以建模为多智能体环境（多智能体系统），异质性参与者——从零售交易者到机构算法——在共同的协议和约束下进行交互（Vidler 和 Walsh 2024a；Epstein and Axtell 1996；Epstein and Schneider 2010；Hull 2022）。不同市场细分领域中智能体的异质性程度各不相同，从而影响系统性行为和稳定性（Vidler 和 Walsh 2024a）。

我们认为，金融市场中的智能体（Vidler 2025）与群体中的独立学习者进行多智能体强化学习（Kaminka 2025）之间存在理论和结构上的相似之处。这些相似之处可为理解经济环境中的分布式决策提供基础。

**受监管的交互与隐式协调**。 与工程化多智能体系统不同，金融系统因法律禁止合谋和信息共享而限制显式协作（Federal Reserve Bank of New York 2021；Bank of England 2015）。尽管如此，协调仍可能通过策略性信号传递、羊群行为以及对共享市场数据的反应等间接渠道隐式涌现，这与群体系统中观察到的印迹式交互相似。因此，这些市场如同去中心化的群体系统，协调在其中无需中央控制便会产生。这一类比促使我们将金融市场视为能够处理信息并动态适应的分布式系统；下文将进一步阐述这一观点。

**分布式信息处理**。 金融市场通过分布式交易机制汇总私人信息，个体信念通过价格信号（股票市场）和可用流动性（债券市场）得以表达。这类似于群体中的分布式感知。在某些市场，例如政府债券市场，这一过程会影响流动性（Pinter 2023；Pinter, Wang, and Zou 2024；Vidler 2025）。系统级反馈凸显了系统中的矛盾：市场（系统）能够通过反馈进行适应，但这种适应能力在平稳、畅通无阻的环境中效果最佳。当面临重大挑战或阻力时，系统在适应的同时维持稳定运行的能力会受到限制。过去几年，仅英国和美国的债券市场就发生了多次系统性冲击（（Pinter 2023），（Duffie 2020））。我们的研究部分受到这些挑战的推动，也关注在全局目标至关重要（即债券市场正常运行），但个体智能体既追求自身利益、又无法明确协调以实现全局结果时，连接各个市场的可能性。下文将介绍与债券市场流动性直接相关的文献。

### 市场流动性：政府债券市场的双边市场结构与流动性供给

双边场外交易（OTC）市场中的流动性，源自交易对手之间的去中心化交互，而非集中式订单簿。许多机构市场（包括政府债券市场）都集中进行资产交易。例如，在英国政府债券市场（即英国金边债券市场）中，债券通过受监管交易商（GEMMs）网络发行和交易[1],；澳大利亚、美国和加拿大也有大体相似的安排（Cheshire 2015）。尽管各市场的制度细节有所不同，但这类市场中的大型投资者通常通过这些中介机构（或其职能等同机构）进行交易。大型参与者提供报价并承接风险，但同时面临资产负债表、资本和监管约束（Pinter, Wang, and Zou 2024; Czech and Pinter 2022）。

在双边场外交易环境中，交易价格往往以公开基准或参考利率为锚定，而非完全由每一笔本地撮合决定。流动性约束源于库存、融资能力以及交易关系网络结构所带来的限制（Duffie and Singleton 2015; Frank J. Fabozzi 2005; Pinter 2023）。因此，流动性是众多去中心化决策（例如库存管理和风险承受能力）共同产生的结果，而不是由中央控制的投入（Frank J. Fabozzi 2005）。监管机构和监测机构（英格兰银行2015）会观察清算成交量，即转化为已执行交易的债券数量；这一指标用于评估**市场流动性**。本文将交替使用“市场流动性”和“流动性”这两个说法。

关于如何将流动性建模为一种由主体行为涌现出的全局结果，仍存在空白。现有方法要么刻画市场微观结构的特征，要么关注参与者的策略互动。我们

[1]https://www.dmo.gov.uk/responsibilities/gilt-market/market-参与者/

通过提出一种模型来弥补这一空白，该模型将流动性博弈（形式化描述双边场外交易的约束与激励机制（Vidler and Walsh 2024b））与基于理性群体的去中心化多智能体学习（Kaminka 2025）相结合。这一统一框架使智能体能够依据局部奖励独立行动，同时通过涌现式清算（债券交易）共同再现系统级流动性。我们进一步允许智能体“学习”如何交易更大规模的债券，从而提升系统整体流动性。由此，该框架将微观层面的决策过程与宏观层面市场结果联系起来，为研究流动性的形成提供了一种无需集中式协调和智能体策略学习的新方法。

### 应用于金融多智能体系统的博弈论

**博弈论**。博弈论方法已应用于金融市场，但大多数框架分析的是**微观层面交互**，而非整体市场行为，因此难以洞察独立交易智能体大规模系统的涌现属性，例如流动性分析。核心挑战在于扩展博弈论见解，以捕捉多参与者市场中复杂的相互依赖关系。

在此背景下，异质和同质多智能体系统的应用涵盖小规模和大规模场景，尤其涉及博弈论、谈判和公平性（Wooldridge 2009；Axtell and Farmer 2022；Parkes 2006；Shoham and Leyton-Brown 2008；Fatima, Kraus, and Wooldridge 2014），相关研究还包括（Fagiolo et al. 2017；Lussange et al. 2020；Vermeir and Bersini 2015）。智能体推理与通信的基础为合作提供了基于逻辑的框架（Fatima, Kraus, and Wooldridge 2014；Kl¨ugl, Fehler, and Herrler 2005年；Thangarajah, Padgham, and Harland 2002）。自动化谈判和不确定性下的决策方面的扩展融合了博弈论与人工智能（Baumeister 2019），并取得了有关重复交互和“玩错博弈”界限的研究成果（Burkov and Chaib-Draa 2014；Meir and Parkes 2018年）。ABM中的目标建模研究包括（Winikoff et al. 2002），而模拟被认为是再现复杂动力学的关键（Bai, Raskob, and M¨uller 2020年）。这些理念为自主竞价智能体（包括金融智能体）以及（Wellman, Greenwald, and Stone 2007；Lin et al. 2014；Shoham 2008）中探讨的面向市场的多智能体系统奠定了基础。我们利用这些要素并扩展LG，使其具备多智能体学习能力，旨在让市场参与者（智能体）学会采取最佳交易策略，以最大化自身流动性（出于利己目的），同时整体上受益于市场流动性的最大化。

**马尔可夫博弈与市场记忆**。马尔可夫决策过程和随机（马尔可夫）博弈因其数学上的可处理性，以及对无记忆状态转移的假设，在金融建模中得到了广泛应用（Shapley 1953年；Littman 1994年）。马尔可夫性质假设，未来的市场状态仅取决于当前状况，而与历史轨迹无关——然而，随着实证证据显示，收益、波动率和流动性（订单流）具有长期记忆和路径依赖行为，这一假设正受到越来越多的质疑（Ding、Granger 和 Engle 1993年；Lillo 和 Farmer 2004年、2005年）。尤其是，波动率动态表现出偏离马尔可夫模型的粗糙（分数阶）行为（Gatheral、Jaisson 和 Rosenbaum 2018年）。同样，交易系统由重复博弈构成，参与者认为了解与之交易的交易对手的特征具有价值（Massa 和 Simonov 2003年）。

在这些方法中，仍存在一个缺口：现有模型无法将局部智能体激励（无论是否进行学习）与全局流动性的涌现这一结果统一起来。

我们通过提出一个统一框架来弥补这一缺口，该框架结合了流动性博弈（Vidler and Walsh 2024b）的严谨结构，以及通过理性群体实现的去中心化学习（Kaminka 2025）。我们表明，智能体能够在局部进行学习，同时保留合法非协调，从而构建出一种可扩展、能自适应的模型，解释流动性如何在复杂金融系统中内生涌现。

### 金融市场的群体特征

#### 奖励机制与群体

多智能体金融系统中的奖励机制涵盖个体激励与系统层面目标（Wooldridge and Jennings 1995；Shoham 2008年）。在智能体层面，奖励自然对应于交易带来的货币收益。在集体层面，奖励则对应于市场属性，例如稳定性、流动性和持续参与。在某些市场（如债券市场）中，智能体直接受益于流动性（能够在希望的时间交易想要交易的债券），而非方向性价格动态。

本文旨在统一金融研究与群体研究的两个方面：我们在马尔可夫团队博弈形式体系中，构建了一个用于群体规模市场流动性的统一框架，将 LG 的方法（收益与聚合流动性挂钩（Vidler 和 Walsh 2024b））

与理性群体相结合（通过差分奖励使自利型智能体保持一致（Kaminka 2025；Tumer 和 Agogino 2007））。我们明确指出，差分奖励是智能体对全局流动性目标的边际贡献，从而提供一种透明且激励相容的信号，在保持智能体独立性的同时避免协调或合谋。这将关注点从双方交易转向大规模群体中涌现的动态，将个体交易激励与整体市场流动性等系统层面属性联系起来。

## 模型与方法描述

本节详细介绍理论模型的定义。以下小节将说明环境（随机配对、清算规则）、智能体类型和学习信号。所有内容均以双边交易市场为背景，其中团队目标是总流动性 $G$（撮合总成交量）。

## 流动性博弈

流动性博弈由 Vidler 和 Walsh（2024b）提出，其定义为“一种由两个智能体（或玩家）参与的非合作同时行动博弈，用于模拟双边市场机制”。该博弈的独特之处在于收益是内生的。

**定义 1**（流动性博弈）。 博弈实例是一个元组$I = (N, B, U)$, 定义如下:

- **玩家**。 $N = \{i, j\}$ 是玩家集合，其中 $|N| = 2$。
- **债券余额**。 每位玩家 $i \in N$ 持有私人债券余额

$$B_i \in \mathbb{Z} \quad \text{with} \quad x \leq B_i \leq y\,,$$

其中 $x, y \in \mathbb{Z}$ 是公开市场参数。通常$x = 0$且 $y = c$, 其中 $c$是给定债券的总发行量（公开可得）。

- **行动**。 玩家 $i$可采取的行动集合为

$$A_i = \{0, 1, \ldots, |B_i|\}\,,$$

并且，动作 $a_i \in A_i$ 表示玩家 $i$在一回合中交易的债券数量。等价地,

$$0 \leq a_i \leq |B_i|\,.$$

- **效用**。玩家 $i$ 的效用（奖励）函数 $u_i$： $A_i \to \mathbb{Z}$ 为

$$u_i(a_i; B_i) = |a_i| \quad \text{for all } a_i \in A_i.$$

## 博弈动机：重复阶段博弈

我们将这一过程建模为重复的（单次）阶段博弈：每个回合开始时，博弈状态都会重置；每个回合中会进行多场双边智能体博弈，使每个智能体每回合仅有一次交易机会。在两个智能体之间的阶段博弈中，双方不使用价格，而是同时各自采取（可能不同的）策略，私下选择一笔数量为 $a_i \in B_i$、以 $0 \leq |a_i| \leq |B_i|$ 进行交易的债券，力求使其债券余额趋近于零但不低于零，同时隐瞒真实库存，以保留信息优势并满足监管规范。选择可能基于对交易对手的信念；智能体代表人类玩家，而人类玩家通常会寻求一次性完成全部清算。我们采用两种博弈机制：只有当双方提交的数量彼此都可接受时，交易才会执行；另一种是“最小成交”机制，即按最小数量成交。虽然不属于本文的研究范围，但我们指出，这些玩家偏好可以纳入更复杂、更丰富的博弈设定（例如惩罚方法）中，并将其留待未来研究。

在 (Vidler and Walsh 2024b) 的经典博弈中，一次性同时行动变体的收益和福利设定为：一名卖方玩家拥有 $B_i > 0$，一名买方玩家拥有 $B_j < 0$，双方各自选择一个非负行动 $a_k \in \mathbb{Z}_{\geq 0}$，且 $a_k \leq |B_k|$。只有在正数精确匹配时才会成交： $q = a_i$ 若$a_i = a_j > 0$，否则为 $q = 0$。每位玩家的收益为$q$。这体现了共同目标，即将余额移向零；任何不匹配（或提交零）都不会改变头寸。

智能体可以属于两个互斥群体之一，即大型或小型企业。

设有一个由智能体 $N = \{1, \ldots, n\}$ 构成的有限集合。每个智能体 $i \in N$ 属于以下两种互斥类型之一:

$$\theta_i \in \Theta = \{\text{large}, \text{small}\}.$$

类型 $\theta_i$ 表示智能体 $i$ 的企业规模。企业规模会影响任何交易中债券收益分布的均值，但不会影响其尾部行为。

我们从两个方面扩展 LG，以提升其适用性的丰富程度:

**清算规则：** 若一笔交易以 $|a_i|$ 和 $|a_j|$ 中较低者为准，从而使双方交易的流动性最小，则称该交易已发生。我们将此变体称为“最小成交”变体。这也自然地将差异奖励的概念扩展为每个智能体期望交易的数量与实际成交数量之间的价值差。形式化地说

我们可以看到，定义 1 中，市场中至少可能存在两种策略:

1. **LG-Exact（基线）**。仅当 $a_i = a_j > 0$ 时才发生交易。交易数量为 $q = a_i$，且 $u_i(q) = q$。 2. **LG-MinFill（扩展）**。当 $a_i > 0$ 和 $a_j > 0$ 时发生交易，成交数量为 $q = \min(a_i,\ a_j)$。收益为 $u_i(q) = q$。

除非另有说明，基准模型的均衡结论均指 LG-Exact，而标记为“MinFill”的学习实验则采用第二条规则。需要指出的是，在 LG-Exact 中，每个对角策略组合 $(a,\ a)$，只要 $a \in \{1, \ldots, \min(|B_i|,\ |B_j|)\}$，都是纯纳什均衡（且 (0, 0) 也是均衡）。在 LG-MinFill 中，唯一的纯纳什均衡是 $a_i^\star = a_j^\star = \min(|B_i|,\ |B_j|)$。

将 LG 扩展到两种机制时，我们把模型推广到严格约束之外。

## 学习型智能体

我们将博弈扩展为重复阶段博弈：每个回合开始时重置智能体的初始余额，但智能体可以学习尝试交易的最优债券数量。具体而言，智能体试图学习应提供多少债券，才能使交易量最大化。学习时需要将交易对手所属的智能体类型的概率作为背景信息（这会影响可能的交易数量），同时还要力求使交易产生的总流动性最大化。

我们采用 Tumer 和 Agogino（2007）提出的差分奖励作为每个智能体的奖励信号。我们的研究大量借鉴了 Kaminka 2025 的工作，该工作展示了合理的去中心化群体智能体在提供环境层面效益方面的强大能力。对于智能体 $i$，其联合结果为$\mathbf{z}$，反事实结果为$\mathbf{z}^{-i}$（即不包含 $i$ 的系统），差分奖励衡量该智能体对全局目标的边际贡献 $G$，定义如下:

$$D_i \;\equiv\; G(\mathbf{z}) - G(\mathbf{z}^{-i}) \tag{1}$$

在我们的设定中， $G$是一种流动性指标，因此 $D_i$ 量化了智能体 $i$ 在保持智能体独立性并避免显式协调的同时，如何改善市场流动性。差分奖励衡量的是智能体的存在对整体系统性能的边际贡献

。例如，这种边际贡献可以通过总交易量等指标来量化。这样，在双边交易环境（例如债券市场）中应用差分奖励（Vidler and Walsh 2024b），就能建立一个框架，评估纳入或排除智能体所产生的非线性影响；该框架不再局限于简单的按比例分摊贡献，还能捕捉由智能体交互产生的涌现效应。我们建议采用这种方法来理解市场流动性。

将流动性博弈（内生 $G$）与理性群体（差分奖励）相结合，既能使个体理性学习与市场层面的目标保持一致，又无需显式协调。我们注意到，这种"边际贡献"视角在概念上类似于风险归因中的常见理念（例如，对投资组合风险的边际或增量贡献）（Hull 2018）。

## 统一金融群体模型的实验

本节详细介绍我们用于测试程序化模型的环境。代码库实现了一个多智能体仿真环境，用于研究简化市场环境中的协调与流动性供给。我们使用该仿真环境探索不同清算规则和奖励机制对智能体行为及市场总体结果的影响。

### 仿真环境

仿真的核心是 LiquidityGameEnv 类，它对一个由固定数量智能体（n agents）组成的市场进行建模。在每个回合¯（时间步）中，智能体会被随机配对。每对智能体同时向对方宣布自己愿意提供用于交易的流动性数量。这些信息绝不会广播给其他智能体。实际交易量取决于所选的撮合规则。交易完成后，智能体的余额会更新，随后开启新一轮回合，此时智能体的余额可能已经发生变化。根据初始余额范围（小型智能体的最大余额和大型智能体的最大余额），将智能体划分为小型 或 大型 ，以便研究非对称流动性禀赋。交互模型示意图见图1。 ¯

**模拟设置**。 我们使用现实市场中的估计智能体数量（Pinter 2023），将智能体数量设为1300，其中大型和小型参与者分别占33%和66%（参考 Vidler 和 Walsh 2024b 中的近似值）。分析运行10,000个回合。每个回合中，各智能体的初始余额上限设为：小型 $\leq 10$ 单位；大型 $\leq 40$ 单位。智能体在每个回合开始时，根据其属于大型或小型智能体群体的概率，在1与相应上限之间随机初始化余额。我们的目标仍然是测试智能体学习应提供多少债券进行交易的能力，从而最大化其交易量以及整个环境的流动性。

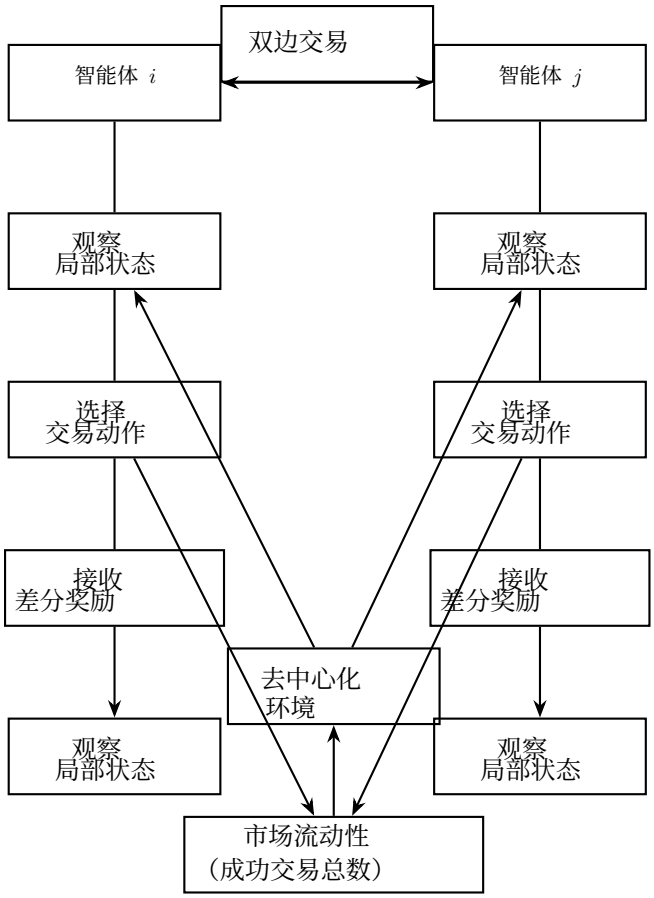



图1：流动性群体模型示意图。智能体$i$ 和 $j$ 通过双边交易进行交互。行动影响全局流动性，并通过差分奖励转化为局部奖励信号。

### 学习型智能体（表格型Q学习）

我们使用简单的Q学习智能体。然而，该模拟比较了采用不同奖励机制的学习型智能体与非学习型基线智能体的表现。所有学习型智能体共享相同的表格型Q学习核心，唯一的区别在于其奖励信号 （差异型、局部型或全局型）。

**Q表**。 状态 $s$ 是智能体当前的余额；动作集合为 $A(s) = \{1, \ldots, s\}$（提供的金额）。Q表中每个离散余额对应一行，每个可行动作对应一列。

**更新**。 执行 $s$ 中的 $a$并观察到奖励 $r$ 和下一个状态 $s'$后，

$$Q(s,a) \leftarrow Q(s,a) + \alpha\Big[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\Big],$$

固定 $\alpha = 0.1$。奖励项 $r$ 由所选信号实例化；不同信号之间没有其他超参数差异。

**行动选择（ε-贪心）**。智能体以 $\epsilon$ =0.2 的概率通过从 $A(s)$ 中均匀采样来探索；

否则，它会进行利用，选择任意 $a \in \arg\max_{a'} Q(s, a')$（若有并列，则随机均匀选择）。

学习型智能体的区别在于每个步骤结束后，它们收到的信号所采用的**奖励模式**。

1. **差分奖励：**智能体获得差分奖励，用以近似衡量其对系统流动性的边际贡献。某个智能体的奖励等于其配对中的交易数量，减去在该智能体报价为零（而其伙伴的报价保持不变）时本应达成的交易数量。在最小成交量规则下，这可简化为交易数量 $q$。

2. **局部奖励：**智能体根据局部效用函数获得奖励：$R_i = 2q - a_i - \pi_{ij}$，其中 $q$ 为交易数量，$a_i$ 为该智能体的报价，$\pi_{ij}$ 为重复选择上一步伙伴所产生的惩罚。该惩罚旨在探索逆向选择，设定为适度的 0.1 个单位。

3.**全局奖励：** 智能体获得该步骤中市场所有配对达成的总流动性。这要求智能体学会为全局结果作出贡献，却无法直接获得关于自身影响的反馈。

**基线智能体**实现了两种非学习型基线策略以供比较：

1. **随机智能体：**每一步中，随机智能体都会提供一个均匀随机的数量，范围从 1 到其当前余额（含两端）。

2. **贪婪智能体：**每一步中，贪婪智能体都会将其当前全部余额用于交易。该策略旨在通过“贪婪”来最大化个人交易量。根据最小成交量规则，如果贪婪智能体提供的数量多于其交易伙伴，则会对记录的流动性施加惩罚，以反映过度提供可能产生的成本；惩罚金额设为超额提供数量的 20%。这样设计是为了惩罚智能体纯粹的贪婪行为。

## 指标

模拟会逐回合跟踪若干指标，以评估不同配置的表现：智能体类型（学习型、非学习型及不同的学习方法）和撮合规则（精确撮合和最小成交量规则）。

- **各回合累计流动性 $G(z)$：**给定回合中所有交易对的交易数量总和（见图3和图2）。

- **成交率收敛：**观察每回合中智能体之间交易的余额占各回合初始余额的平滑百分比。（见图5）。

- **命中率：**成功与 $q > 0$ 完成交易的配对智能体所占比例（见图4）。

通过在不同清算规则、奖励模式和基线策略下开展实验，并配置异质性，该代码库能够全面分析影响这一去中心化模拟市场环境中流动性的因素。

## 结果

实验结果证明了差分奖励学习框架在不同市场清算机制下的有效性（见图2）。在允许部分订单执行的“最小成交量”撮合规则下，差分奖励方法实现了最高的清算总流动性，优于替代学习方法（“学习（全局）”和“学习（局部）”）以及基线智能体。与“精确撮合”相比，这一优势尤为明显：差分奖励方法仍保持强劲表现，而其他学习方法则收敛到相近水平，但仍优于基线基准，如图5和图6所示。

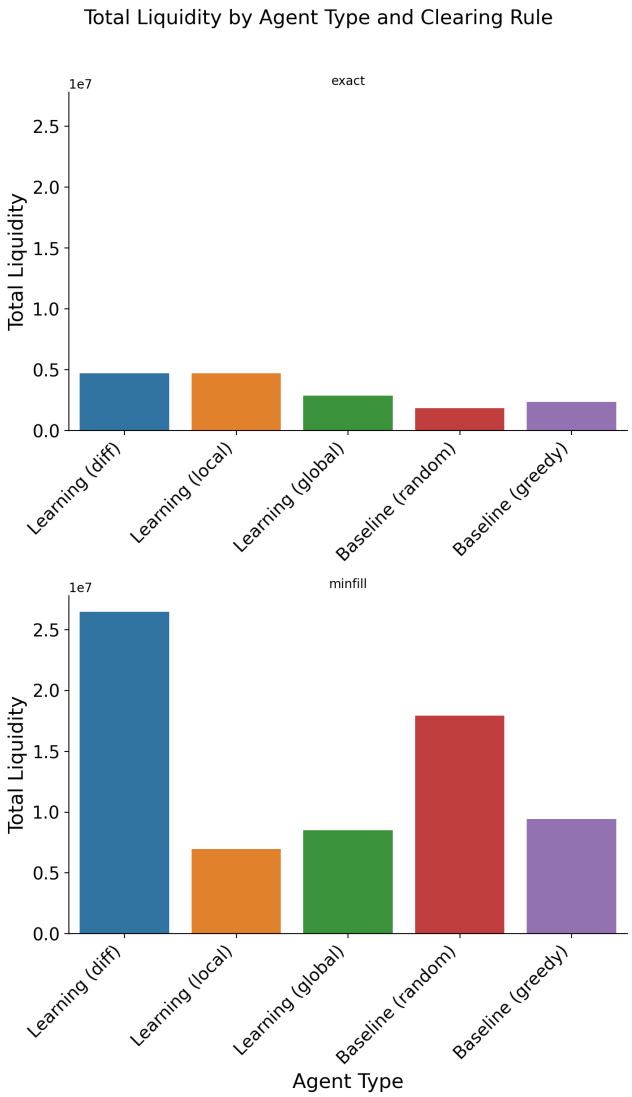



图2：学习型智能体超越基线方法。

累计流动性轨迹进一步印证了差异奖励学习的优势，尤其是在“最小成交”模式下，该框架有效分配功劳的能力

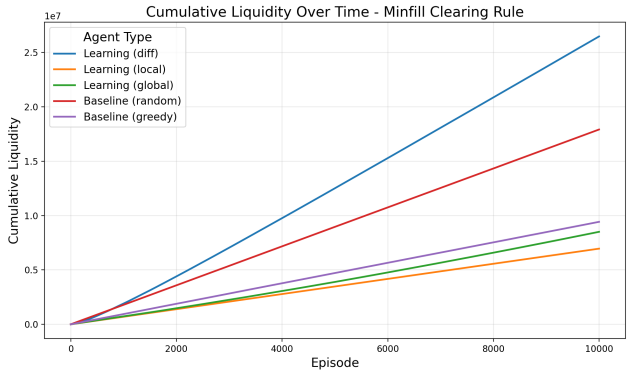



图3：MinFill规则下各回合的累计流动性

能转化为跨回合更高的流动性累积（图3）。具体而言，每回合已清算金额占初始余额的平滑百分比显示，在“最小成交”模式下，差异奖励学习始终实现最高的清算效率，展现了其相较于可用资源最大化流动性利用率的能力。




图4：学习算法的命中率很高，此处展示的是精确成交规则下的结果（注意：根据构造方式，MinFill下的命中率为100%）。**注意：**学习（$exact_diff$）和学习（$exact_local$）的结果几乎完全相同，图表相互重叠

正如预期，允许部分清算的机制产生了更高水平的流动性；不过，在精确成交环境中使用差分奖励的学习型智能体，尽管受到限制，仍能在70%的情况下成功交易（图4）。这些发现表明，在微观结构下，差分奖励是为流动性供给提供实用且激励相容的信号。

虽然“精确撮合”规则限制了整体表现并缩小了不同方法之间的差距，但差分奖励框架仍具有竞争力，这凸显了它在不同市场微观结构下的稳健性，以及其作为金融市场模拟中多智能体协调的强大方法的潜力。

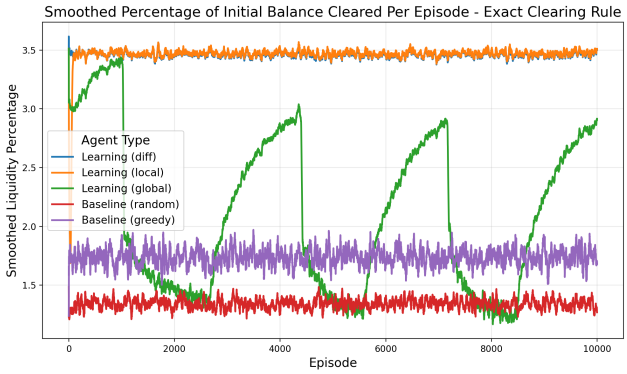



图5：清算率在学习过程中明显较早收敛；精确规则下的差异函数与局部学习完全相同。

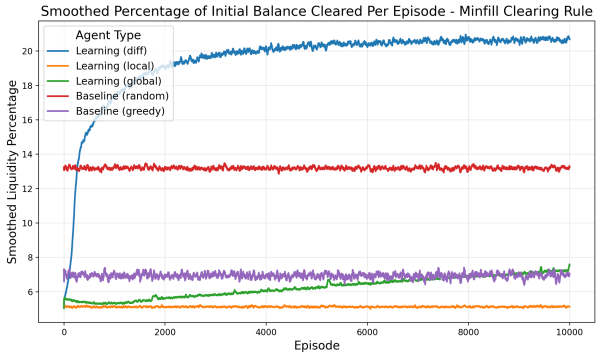



图6：MinFill学习的清算率。

## 结论

在本研究中，我们展示了如何将流动性博弈与去中心化差异奖励学习（理性群体）相结合，以检验自利型智能体能否提升自身和市场的流动性。结果支持了这两项假设。**H1：** 在部分清算规则（最小成交）下，差分奖励智能体实现了最高的总流动性和单个智能体吞吐量，表现优于其他局部/全局信号和非学习基线方法。**H2：**即使在精确成交约束下，尽管表现差距有所缩小，差分奖励智能体也能迅速适应，在训练早期便达到 $\sim$ 70% 的交易成功率，并在此后持续保持。这些发现表明，在不同的市场微观结构下，差分奖励都能为流动性供给提供实用且激励相容的信号；尤其是在允许部分清算时，效果提升显著。这些结果对于不接受智能体协调的去中心化系统尤其有用。差分奖励实现了边际贡献效用：智能体优化与其对总流动性的贡献相对应的私人收益，从而促进双边市场中改善福利的结果。

本文表明，将流动性博弈与理性群体相结合，可以设计出局部激励，

使短视的最佳响应带来系统层面的收益，并将流动性外部性转化为个体理性目标。这为可扩展的市场设计提供了一种实用范式，无需中央协调者或智能体之间的协调。

**局限性与后续工作**。我们的实验采用了风格化动态（随机配对、重置禀赋、不形成价格），因此评估其外部有效性时需谨慎。未来工作将（i）加入持久状态和摩擦（库存、交易成本），（ii）测试更丰富的清算机制和市场冲击模型，以及（iii）分析智能体异质性和部分可观测性的敏感性。将分析扩展到真实数据或基于智能体的模型与强化学习设置，可以进一步探究差分奖励的扩展能力。

## 参考文献


Axtell, R. L.; and Farmer, J. D. 2022. 经济学与金融学中的基于智能体的建模：过去、现在与未来。《经济文献杂志》, 1–96. Bai, S.; Raskob, W.; and M¨uller, T. 2020年. 基于智能体的模型 . 《辐射防护》 , 55. 英格兰银行 . 2015. GEMM 指南：英国政府债券市场中债务管理办公室和一级交易商的职责指南. Baumeister, D. 2019. 决策的未来, 15–20. 施普林格国际出版公司: 施普林格国际出版公司. ISBN 978-3-030-18050-8. Burkov, A.; and Chaib-Draa, B. 2014. 多智能体系统的重复博弈：综述. 《知识工程评论》, 29(1): 1–30. Cheshire, J. 2015. 债券市场中的做市. 澳大利亚储备银行公报 *2015*年*3*月季刊, March 2015(March): 63–74. Czech, R.; and Pinter, G. 2022. 公司债券市场中知情交易与客户与交易商关系的动态. 英格兰银行, 工作人员研究论文(895). Ding, Z.; Granger, C. W. J.; and Engle, R. F. 1993. 股票市场收益的长记忆特性与一个新模型. 《实证金融学杂志》, 1(1): 83–106. Duffie, D. 2020. 仍是世界安全避风港吗？哈钦斯工作论文, 62. Duffie, D.; and Singleton, K. J. 2015. 可违约债券期限结构建模. 《金融研究评论》, 12(4): 687–720. Epstein, J. M.; and Axtell, R. L. 1996. 成长中的人工社会：自下而上的社会科学. 麻省理工学院出版社图书. 麻省理工学院出版社. ISBN ARRAY(0x58c9d590). Epstein, L.; and Schneider, M. 2010. 模糊性与资产市场. Fagiolo, G.; Guerini, M.; Lamperti, F.; Moneta, A.; and Roventini, A. 2017. 经济学与金融学中基于智能体模型的验证. 仿真基础、方法与应用. Fatima, S.; Kraus, S.; and Wooldridge, M. 2014. 自动化谈判原理. 剑桥大学出版社. ISBN 9781107002548.

纽约联邦储备银行。2021年。《全市场中央清算的净额结算效率》。Frank J. Fabozzi, S. M. 2005年。《固定收益证券手册》第七版。纽约：麦格劳-希尔。Gatheral, J.; Jaisson, T.; and Rosenbaum, M. 2018年。波动率是粗糙的。《定量金融》，18(6): 933–949。Hull, J. C. 2018年。《风险管理与金融机构》。新泽西州霍博肯：约翰·威立父子公司，第5版。ISBN 978-1119448112。Hull, J. C. 2022年。《期权、期货及其他衍生品》。伦敦：培生，第11版。ISBN 9781292410654。Kaminka, G. A. 2025年。群体可以是理性的。《英国皇家学会哲学汇刊*A*辑：数学、物理与工程科学》，383(2289：专题“群体系统的未来之路”，由Hussein A. Abbass和Sanaz Mostaghim编纂并编辑）。Kl¨ugl, F.; Fehler, M.; and Herrler, R. 2005年。环境在多智能体仿真中的作用。见Weyns, D.; Van Dyke Parunak, H.; and Michel, F.编，《多智能体系统的环境》，127–149。柏林、海德堡：施普林格·柏林·海德堡。ISBN 978-3-540-32259-7。Lillo, F.; and Farmer, J. D. 2004年。有效市场的长记忆。《非线性动力学与计量经济学研究》，8(3)。Lillo, F.; and Farmer, J. D. 2005年。供需长记忆理论。《物理评论*E*辑》，72(6): 066122。Lin, R.; Kraus, S.; Baarslag, T.; Tykhonov, D.; Hindriks, K.; and Jonker, C. M. 2014年。Genius：支持通用自动谈判代理设计的集成环境。《计算智能》，30(1): 48–70。Littman, M. L. 1994年。马尔可夫博弈：多智能体强化学习的框架。见国际机器学习大会，157–163。Lussange, J.; Belianin, A.; Bourgeois-Gironde, S.; and Gutkin, B. 2020年。金融市场中的学习与认知：基于智能体模型的范式转变。*IntelliSys*。Massa, M.; a nd Simonov, A. 2003年。声誉与交易商间交易：美国国债市场的微观结构分析。《金融市场杂志》，6(2): 99–141。Meir, R.; and Parkes, D. 2018年。玩错博弈：界定多样化智能体群体中的外部性。arXiv:1411.1751。Parkes, D. C. 2006年。《迭代组合拍卖》第*2*章：迭代组合拍卖。麻省理工学院出版社。Pinter, G. 2023年。2022年英国金边债券市场危机剖析。英格兰银行，工作人员工作论文1019。Pinter, G.; Wang, C.; and Zou, J. 2024年。规模折价与规模惩罚：债券市场的交易成本。《金融研究评论》，37(7): 2156–2190。Shapley, L. S. 1953年。随机博弈。《美国国家科学院院刊》，39(10): 1095–1100。Shoham, Y. 2008年。计算机科学与博弈论。《*ACM*通讯》，51(8): 74–79。Shoham, Y.; and Leyton-Brown, K. 2008年。《多智能体系统：算法、博弈论与逻辑基础》。美国纽约：剑桥大学出版社。

Thangarajah, J.; Padgham, L.; and Harland, J. 2002年。BDI智能体中目标的表征与推理。《澳大利亚计算机科学通讯》，24(1): 259–265。Tumer, K.; and Agogino, A. 2007。分布式基于智能体的空中交通流量管理。见第六届自主智能体与多智能体系统国际联合会议论文集（*AAMAS '07*），255。Vermeir, A.; and Bersini, H. 2015。经济学与金融领域基于智能体模型编程的最佳实践。《经济学与数学系统讲义》。Vidler, A. 2025年。《金融市场中的智能体智能：用于双边固定收益交易、嵌入大语言模型的多智能体系统》。博士论文，新南威尔士大学，澳大利亚。Vidler, A.; and Walsh, T. 2024a。解读场外政府债券市场流动性：市场动态的基于智能体模型。见即将出版，*2025*年*IEEE*金融工程与经济计算智能研讨会。Vidler, A.; and Walsh, T. 2024b。非合作流动性博弈及其在债券市场交易中的应用。arXiv:2405.02865。Wellman, M.; Greenwald, A.; and Stone, P. 2007。《自主竞价智能体：来自交易智能体竞赛的策略与经验》。麻省理工学院出版社。ISBN 978-0-262-23260-9。Winikoff, M.; Padgham, L.; Harland, J.; and Thangarajah, J. 2002。智能体系统中的声明式与程序式目标。见第八届知识表示与推理原理国际会议论文集，KR'02，470–481。美国加利福尼亚州旧金山：摩根考夫曼出版社。ISBN 1558605541。Wooldridge, M. 2009。《多智能体系统导论，第二版》。美国新泽西州：约翰·威立父子公司。Wooldridge, M.; and Jennings, N. R.编，1995。《智能体：*ECAI-94*智能体理论、架构与语言研讨会》，第890卷，收入《计算机科学讲义》。荷兰阿姆斯特丹：施普林格。