It's Rational for AI Agents to Procrastinate

Nico Schiavone [1 2] **Eldan Cohen** [3] **Sheila McIlraith** [1 2]

Abstract

We propose labour games, a class of unmediated mixed-motive cooperation games where rational agents must collectively complete a task by taking a negative-reward labour action U times over T timesteps. Each agent aims to maximize their utility by minimizing their contribution, achieving an optimal balance of labour and completion chance. We first study a single agent in a system of opaque agents, and find a tendency for delaying labour actions, dubbed critical completion, similar to procrastination. Using this result, we investigate the multi-agent system, defining the rules governing the general behaviour of agents in a labour game. We find that agents have a preference for contiguous action leading to critical completion and characterize this behaviour as the state of commitment which we show is common knowledge. We show that commitment prevents less capable rational agents from contributing in labour games and illustrate the link between commitment and emergent communication. With these principles as a guide, we propose a mechanistic mitigation to the undesirable phenomena and analyze its effects on the multi-agent system. We experimentally illustrate the discoveries of our theoretical analysis using a multi-agent reinforcement learning encoding of the game. Finally, we provide a discussion on the implications of our results, the inefficacy of current reward structures for practical multi-agent systems, and the open question of how to build appropriate incentive structures for productive agents.

[1]Department of Computer Science, University of Toronto, Ontario, Canada [2]Vector Institute for Artificial Intelligence, Canada [3]Department of Mechanical and Industrial Engineering, University of Toronto, Ontario, Canada. Correspondence to: Nico Schiavone <nqs@cs.toronto.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

1 Introduction

AI agents are fast emerging as a tool to alleviate the burden of many recurring and labourious tasks for businesses and individuals alike. However, when multiple AI agents interact without oversight, the outcomes can be volatile and unexpected. As a result, systems where such volatile behavior is possible are typically mediated by a human, limiting their real-world feasibility and scalability. As a case in point, when agents' actions are in service of *chores*, those actions are disincentivized, necessitating mediation to reliably solve (Guo et al., 2023). Understanding the underlying interaction dynamics of rational, self-interested agents is key to their safe implementation, and motivates much of the recent work on cooperative AI (e.g. (Dafoe et al., 2020; Clifton and Riché, 2020; Yu et al., 2022)) and the work in this paper.

A common cooperative dilemma is division of labour, where groups of individuals working to a collective goal split up (burdensome) tasks. Division of labour is a phenomenon found in societies of insects, microorganisms, and other mammals, and the efficient solution to this dilemma is often cited as a primary driver behind evolutionary success (Page, Robert E. and Mitchell, Sandra D., 1998). Occasionally, division of labour is as simple as asymmetric individuals performing their most suited tasks; however, when two or more individuals work on the same task, they often fail to achieve the collective goal, presenting significant challenges when unmediated (Liu and Zhang, 2019).

The most common example of this is in human group work, where all parties prefer to see the goal completed, but often would rather perform other tasks than contribute to the common goal. This results in a game of chicken, where one member (usually with low risk tolerance) is exploited by the others and performs the majority of the work (Rapoport and Chammah, 1966). Thus, many wish for an altruistic goal-minded third-party supervisor that can force an even split of work between the group members, but this is rarely available.

In multi-agent systems, an equivalent problem is that of automated house cleaning, which is often performed by systems of independently acting robots (Memmesheimer et al., 2024). The cleaning tasks are represented as chores, which incur a negative penalty when performed, but there

is incentive for successfully cleaning the whole space. This multi-agent system exhibits the same issues in that no robot is directly incentivized to clean. If the problem is modified such that the cleaning agents are rewarded for labour, then they will act inefficiently and often not complete the desired task at all (Dulac-Arnold et al., 2019). Thus, a third-party controller is almost always used to ensure a near optimal outcome while the unsupervised approaches remain largely unexplored (Nash, 1951).

In this paper, we introduce and study the **labour game**, an unmediated mixed-motive cooperation problem with self-interested rational agents, where there exists a collective task with a large penalty for non-completion (equivalently, a completion reward), but contributing to this task incurs a small, agent-specific cost. We use this problem as a framing device to characterize the behaviour of rational agents in the presence of a greater system of rational agents, and investigate the emergence of cooperative phenomena.

We explore these key concepts to understand the fundamental problems with incentivizing agents in chore allocation, and to better understand how to design practical multi-agent systems for burdensome tasks. We begin by investigating a labour game with a single agent in isolation, defining their behaviour under this incentive structure. We then analyze how this behaviour changes when the single agent is inserted into a community of similarly incentivized agents. Through this analysis, we propose a ruleset governing the behaviour of the multi-agent system in a labour game, and provide insight into the emergence of **procrastination**, **communication**, and **cooperation** without explicit enabling mechanisms. We also describe behavioural phenomena, such as the stark noncompletion rate of tasks that require the labour of multiple agents. We establish techniques for mitigating undesirable phenomena and promoting cooperation without separate communication, and offer a reasonable, realistic implementation for a physical system.

Our contributions can be summarized as follows:

- We propose labour games, a class of unmediated chore allocation games
- We describe emergent phenomena in the behaviours of rational agents in a labour game, including productive procrastination and communication, and prove key facts about them
- We provide analysis on how to mitigate the undesirable phenomena and mechanistically encourage more cooperative behaviours
- We illustrate our theoretical results using experiments in multi-agent reinforcement learning

Our work provides a theoretical framework for studying unmediated multi-agent systems focused on chores and other burdensome tasks, addressing a gap in current literature and opening an important avenue for the successful implementation of practical multi-agent systems.

2 Related Work

Cooperation and Safety: The emerging field of cooperative AI has received much attention as of late, producing foundational and cautionary works which tout the benefits and risks of cooperative scenarios (Muglich et al., 2022; Baker, 2020; Dafoe et al., 2021; Tan, 1993; Willis et al., 2025). Many agentic systems are analyzed under the assumption of a general controller for all the agents or some degree of transparency and architectural similarity between them (Hughes et al., 2018; Yu et al., 2022; Zhu et al., 2025). Thus, when agents interact with unfamiliar agents, the results can be unpredictable and dangerous (Hammond et al., 2025; Piatti et al., 2024). Understanding the potential of cooperative systems and mitigating their risks is the primary focus of cooperative AI (Conitzer and Oesterheld, 2023; Han et al., 2024). This brings to question problems such as outcome fairness, definitions of cooperative vs. exploitative behaviour, and behavioural preferences. Cooperative AI researchers have recently become concerned with the trustworthy deployment of agentic AI, including multi-agent systems, and how the result of interaction between two unknown agents can be made reliable and safe (Dafoe et al., 2020; Clifton and Riché, 2020; Zhang et al., 2024). One line of work (e.g. (Rahman et al., 2023; Wang et al., 2024)) trains single agents to cooperate with unknown agents with distinct agendas, but this generally takes an optimistic view of the other agents one might encounter, disregarding exploitative or disruptive agents. We re-analyze the fundamental problem of opaque agent interaction to help better define the default behaviour of these rational agent systems to better inform future research on their design. We hope to use the emerging behaviours (e.g. (Guo et al., 2024)) that have been uncovered by recent cooperative AI research as a pathway to study the game theoretic reasons behind them.

Social Dilemmas in Multi-Agent Reinforcement Learning: Hughes et al. (2018) define social dilemmas as games where all of the following hold: mutual cooperation has a higher collective welfare than mutual defection; mutual cooperation is preferable to being exploited; and either mutual defection is preferable to being exploited, or exploiting a cooperator is preferred to mutual cooperation. Existing work investigating social dilemmas in multi-agent reinforcement learning (Mintz and Fu, 2024; Wang et al., 2021; Hughes et al., 2020; Yocum et al., 2023; Lupu and Precup, 2020; Rios et al., 2023; Leibo et al., 2021; Tennant et al., 2023; McKee et al., 2023) seeks to understand and solve social

dilemmas through extrinsic mechanisms or imbuing agents with certain behaviour preferences. For extrinsic mechanisms, Haupt et al. (2024) develop the idea of contracting in multi-agent reinforcement learning as unconditional reward transfer attached to certain actions. Hughes et al. (2020) use a forced-action version of contracting to a similar end, allowing for an agreement and subsequent fixed action. For intrinsic properties, Hughes et al. (2018) investigate using inequity aversion without explicit mechanisms for cooperation, and find that similar results can be achieved this way. There are many examples of both strategies, e.g. (Wang et al., 2021; Lupu and Precup, 2020; McKee et al., 2023; Chen et al., 2024) which find varying levels of success in finding more human solutions.

Fair Allocation Problems: Allocation problems are a well studied class of problems that involve fairly distributing divisible or indivisible goods amongst a group of agents. Recent literature on allocation problems involves allocation of 'chores', or bothersome tasks that do not benefit the agent they are allocated to (Li et al., 2023; Garg et al., 2023). Algorithms for this involve notions of envy-freeness, search efficiency, and constrained solutions, e.g. (Cookson et al., 2024; Yin and Mehta, 2022; Aziz et al., 2023; Bhaskar et al., 2022; Ebadian et al., 2022). A common thread throughout allocation problems is the presence of a powerful third-party that can assign the goods to each agent, and is trusted to do so in a fair way (Guo et al., 2023). There is also a distinct class of problems known as division of labour problems, which occur in nature, and often deal with similar solution techniques to chore allocation (Zhao and Zhang, 2023). In our research, we are focused on a problem very similar to chore allocation, only we remove the need for an altruistic third-party, and provide a system for the agents to self-allocate in a temporally extended manner. This is a critical gap in chore allocation literature, and provides a basis for future works studying the unmediated problem.

3 Problem Formulation

We are motivated by a class of problems that see multi-agent systems working towards a collective goal, comprised of one or more burdensome tasks (chores) which may require the contributions of multiple agents. This set of problems may arise when an agent is completing a task and we wish to help it by adding other agents to the system, or when an agent considers the impact of unknown potential contributors. In addition, agents may be preoccupied, and thus will be idle on certain tasks for that time.

Classically, these multi-agent systems are realized with an altruistic third party working towards the collective goal with fairness in mind (Guo et al., 2023). However, this is often unrealistic and expensive, especially when the number of agents is high, so we would like to use multi-agent systems without constant third-party supervision and predefined problem-specific division of labour. To this end, we consider a class of problems where the system of agents operates without a set chore distribution, and provide a framework for analyzing this novel scenario.

We represent this scenario as a temporally extended division of labour where the multi-agent system must complete a certain number of labour units within the time limit. Following the literature on chores (e.g. (Aziz et al., 2023)), we choose to represent labour actions with a reward penalties. We also impose a reward penalty if the task is incomplete at the time limit to represent incentive to complete the task. Agents are disincentivized to labour, but an incomplete task is often worse, so agents prefer that labour is done by other agents but will labour if required. To increase the realism of the setting, these penalties are agent-specific.

Definition 1 (Labour Games) *A labour game $G = \langle U, T, \boldsymbol{r}, \boldsymbol{f} \rangle$ is an N-agent game where a multi-agent system must complete U burdensome units of labour within T timesteps. $\boldsymbol{r} = (r_0, ..., r_{N-1})$ denotes the agent-specific labour penalties, and $\boldsymbol{f} = (f_0, ..., f_{N-1})$ denotes the agent-specific failure penalties. Each agent acts simultaneously at each timestep, choosing either to labour, receiving their penalty r_i, or to idle. If there are less than U units of labour completed when all timesteps have elapsed, each agent receives their failure penalty f_i. If U or more units of labour are completed before T timesteps have elapsed, the game is won.*

Definition 2 (Markov Labour Game) *A Markov Labour Game $\mathcal{M}$ for a labour game $G = \langle U, T, \boldsymbol{r}, \boldsymbol{f} \rangle$ is a fully observable Markov game $\mathcal{M} = \langle S, s_0, \mathbf{A}, L, \mathbf{R}, \gamma, t_{rem} \rangle$ with $U \in \mathbb{Z}^+$ and $T \in \mathbb{Z}^+$. S is a state space; $s_0 \in S$ is the initial state; $\mathbf{A} = A_0 \times A_1 \times \cdots \times A_{N-1}$ is the space of actions for N agents, where $A_i = \{$idle, labour$\}$; $L : S \times \mathbf{A} \to S$ is a transition function; $\mathbf{R} : S \times \mathbf{A} \to \{0, r_0, f_0\} \times \{0, r_1, f_1\} \times \cdots \times \{0, r_{N-1}, f_{N-1}\}$ is a reward function mapping state-action profiles to reward vectors for the N agents. We denote t_{rem} as the current time remaining, beginning at $t_{rem} = T$; u_i as the number of labour actions taken by agent i; r_i as the penalty for agent i for performing the labour action; and f_i as the loss penalty for agent i. We adopt the convention that the timestep $t_{rem} = 0$ is solely for a final win/loss check, so the last actionable timestep is $t_{rem} = 1$.*

In a labour game, agents may only interact indirectly through the task, there is no additional avenue for communication or interference. In addition, we consider units of labour to be wholly equivalent and unlimited up to the number of units required: if two agents labour in the same timestep, it does not matter which agent receives which unit of labour.

Winning a labour game constitutes the multi-agent system completing U units of labour within T timesteps. *Losing* a labour game constitutes more than one unit of labour remaining uncompleted when $t_{rem} = 0$. The solution of a labour game consists of a labour schedule for each participating agent, dictating for each time step whether the agent should execute an idle action or a labour action. Individual agents may be better or worse off than others depending on the notion of fairness used. An *envy-free solution* consists of no agent taking the labour action more than any other agent. An *equal solution* has the ratio of overall labour utilization in each agent as the same. We also recognize other notions of fairness (e.g. (Guo et al., 2023)) would dictate different fair solutions. Refining the concept of completion, we denote the scenario where the labour game is won after using the entire time limit as *critical completion*.

Definition 3 (Critical Completion) *A labour game is critically completed if the final unit of labour is performed at the final actionable timestep $t_{rem} = 1$.*

4 Analysis

In the analysis of labour games, we assume the multi-agent system is composed of rational, self-interested agents that may be of different architectures but share the labour game incentive structure. All agents must share the collective goal of completing the task, and prefer completion to non-completion (i.e., a failure penalty $f_i < 0$). We denote action sequences that result in a utility less negative than f_i as *preferable*. In addition, the penalties will share a strict hierarchy of magnitudes: $|f_i| > |r_i|$ to ensure no agent prefers failure over a single unit of labour. We also assume that multiple labour actions on a single timestep are not strictly necessary to win ($U < T$), and that t_{rem} counts down from $t_{rem} = T$ to $t_{rem} = 0$.

By the formulation of Hughes et al. (2018), this game classifies as a social dilemma: mutual cooperation has a higher collective welfare than mutual defection, exploiting a cooperator is preferable to cooperating, and mutual cooperation is preferable to being exploited. Thus, the best outcome will not be achieved by self-interested action; in this section, we will characterize the behaviour patterns that result from self-interested action and the outcomes that follow. We also consider the realistic aspect that a labour game solution should minimize the total timesteps taken to completion, beyond just the (possibly artificial) time limit T.

4.1 Single-Agent Behaviour

In this section, we consider a single agent in a labour game, and show the default labour patterns of the single agent when operating uninterrupted.

First, we define labour capacity x_i, a unitless quantity, as the ratio of the failure penalty f_i to the labour penalty r_i, equivalent to the number of labour actions an agent i can perform in a labour pattern before failure is preferable. Also, let H_i be the utility of agent i after the game.

Definition 4 (Labour Capacity) *An agent i in a Markov labour game $\mathcal{M}$ has utility H_i, and labour capacity $x_i = \frac{f_i}{r_i}$ representing the number of labour actions the agent can plan to take before failure is preferable: $u_i = x_i \implies H_i = u_i r_i = f_i$.*

A single agent will prefer a labour pattern if and only if an amount of labour $u_i \leq x_i$ prevents failure; therefore, a single agent will only win a labour game if $U \leq x_i$. This directly leads to the first result on when labour is preferable for a single agent.

Proposition 1 *An agent i in a Markov labour game $\mathcal{M}$ will win the game if and only if $x_i \geq U$, and will otherwise idle rather than labouring.*

Proof Sketch: The full proof, as with all the proofs in this paper, is provided in the supplementary material; we provide a sketch of the intuition in its place. The proof for this proposition follows directly from the maximization of utility; an agent i has essentially the following utility $H_i = \max(r_i U,\ f_i)$. Clearly, the agent will not labour if it does not lead to completion, and if completion has a lower utility than failure, then idling is preferable.

4.2 Multi-Agent Behaviour

We now consider a multi-agent system in a labour game, and show how the presence of other (possibly unknown) agents following the same incentive structure affects the labour patterns of the individual agents. We then analyze the emergent phenomena that occur and propose a set of rules governing the behaviour of single agents in the multi-agent system.

The presence of other, possibly productive, agents in a labour game necessarily changes the behaviour of a single-agent. In the single-agent labour game, all timesteps are equally favourable due to the fixed labour penalty. Adding the presence of other agents also adds the possibility of labour contributions from other agents. A single agent will always act to maximize their utility, and therefore maximize the possible labour contributions from other agents by delaying their own actions. We use this behaviour to introduce the concept of an *action discontinuity*, a point in the game where an agent switches from repeatedly taking the 'idle' action to repeatedly taking the 'labour' action or vice versa (e.g. an isolated unit of labour is two action discontinuities).

Definition 5 (Action Discontinuity) *In a Markov labour*

game $\mathcal{M}$, an action discontinuity is a timestep where an agent i takes a different action from the previous timestep.

Lemma 1 *In a Markov labour game $\mathcal{M}$, an agent i acting in isolation will have at most one action discontinuity.*

If multiple agents take the labour action, this transmits new information to each agent, allowing the labour action to be further delayed. The constant delay of labour leads to the final units of labour always being performed on the last actionable timestep $t_{rem} = 1$, which is a defining behaviour of agents in a labour game.

Theorem 1 *An agent i acting in a Markov labour game $\mathcal{M}$ without knowledge of other agents' incentive structures will critically complete all labour games where $x_i \geq U$, where x_i is the labour capacity defined Definition 4.*

Proof Sketch: A single agent i acting without knowledge of any other agents involved in the collective goal prefers success over failure while minimizing u_i as much as possible to maximize the overall reward at the end of the game. Assuming actions impose time invariant penalties, it follows that in the presence of unknown actors, it is preferable for any given agent to act later than earlier to minimize their amount of labour due to the possibility of intervention. A single agent will therefore idle for the first $T - U$ timesteps, at which point idling more causes unpreferable failure. Thus labour will always begin at $t_{rem} = U$ in the absence of interference. If any labour is performed by an outside source, this will be further pushed back to $t_{rem} = U - \Sigma_j u_j$. The last unit of labour is performed on the last timestep in all cases, verifying the claim.

From Theorem 1, it follows that labour actions are performed in contiguous time intervals with the maximum amount of delay on each interval to allow for the maximum contribution from other agents. This gives rise to another condition on their behaviour, the minimization of action discontinuity.

Lemma 2 *An agent i acting in multi-agent system inside a Markov labour game $\mathcal{M}$ will minimize action discontinuity.*

Using Lemma 2, for an agent i if $x_i > 0$ we can define a time $t_{rem} = t_k$ where the agent will begin labour and continue until $t_{rem} = 0$ in the absence of actions by other agents. This time represents the critical point where the agent now prefers contributing to a critical completion over losing the game. We define this behaviour as *commitment*, and posit that this labour pattern is rational and common knowledge amongst the agents.

Definition 6 (Commitment) *An agent i in a Markov labour game $\mathcal{M}$ is considered committed if $|f_i| > |t_{rem} r_i|$ and $u_i > 0$. Committed agents prefer labouring for the remaining amount of time in the labour game over losing the game.*

We use k_i to denote the number of committed agents excluding agent i. As it is irrational to take a labour action if not committed, agents observed taking labour actions must be committed, which affects the decision making of other agents. Therefore, the behaviour of an agent i is informed by k_i at any given timestep, as it represents a promise of future labour by other agents. Then, it is not rational for i to commit if k_i is such that i is unnecessary or insufficient for critical completion. As k_i has only N possible values across all agents, commitments will occur at a discrete number of points where an agent could possibly contribute to a critical completion. We denote these decision points for as t_{k_i}. The timing of the contiguous labour blocks, and therefore the behaviour of all rational agents in labour games, depends entirely on the times t_{k_i} and the number of currently committed other agents k_i.

Theorem 2 *Let k_i be the number of committed agents, excluding agent i, in the Markov labour game $\mathcal{M}$, and t_{k_i} be the commitment decision times for an agent i. The behaviour of an agent i in $\mathcal{M}$ is governed by the number of committed agents; an uncommitted agent will only commit if they are necessary and sufficient to critically complete the game:*

1. $t_{k_i} = \lceil \frac{U - \Sigma_j u_j}{k_i + 1} \rceil$
2. *If* $(t_{rem} \leq t_{k_i}) \wedge (u_i + t_{rem} r_i < f_i) \wedge (t_{rem} k_i < U - \Sigma_j u_j) \wedge (U - \Sigma_j u_j \leq t_{rem}(k_i + 1))$*, at any point, then agent i will commit*
3. *Agent i will labour iff i is committed and $t_{rem} \leq t_{k_i}$*

Proof Sketch: From Theorem 1, any agent i will critically complete a Markov labour game $\mathcal{M}$ if $x_i \geq U$. The first statement represents the modified threshold for critical completion based on the commitment of the other agents. It is similarly known that agents will maximize their overall reward by minimizing the number of labour actions taken. Thus, a non-committed agent committing is only preferable if the agent i is necessary and sufficient for the completion of the game and it maximizes the overall reward when taking account of the previous and future work required for critical completion, which gives the second statement.

A direct consequence of Theorem 2 is that commitment is always accompanied by a labour action. Therefore, no agent will begin to labour unless $\exists i \mid x_i \geq U$, and this labour will begin at precisely $t_0 = U$. Intuitively, this is explained by the rationality of not wasting effort; the opportunity cost of acting without completing the task is high: agents are shown to act only when there is a guaranteed win.

Thus, for all agents j where $x_j \geq U$, commitment will occur precisely at $t_0 = U$, immediately reducing the threshold to $t_j = \lceil \frac{U-j}{j} \rceil$. All agents will switch to idling until $t_{rem} = t_j$, and then all j committed agents will begin another contiguous labour block to critically complete the labour game. This behaviour is explained by the prior statements; all j agents prefer completion and prefer performing all remaining labour over failure, therefore only one of the j agents is strictly necessary, and this is common knowledge. Therefore, all j agents prefer idling until the point of critical completion (from Theorem 1).

4.3 Rational Communication and Procrastination

In this section, we use the results of Theorem 2 to elucidate the key emergent phenomena: procrastination and communication, and show how these are rational behaviours.

Two results are immediately visible from Theorem 2: first, the decision making process of rational agents in a labour game depends solely on the number of agents that have taken the labour action in at least one timestep; second, the idle action is preferable over the labour action until the point where critical completion is not possible. The first result demonstrates the emergence of the first non-idle action as a form of communication between the agents with $x_i \geq U$, indicating their commitment. However, for agents with $x_i < U$, this communicative action can never occur. We call this *productive communication*, as it only takes place if the action itself is never wasteful. Thus, only a subset of agents can viably communicate and therefore cooperate.

Lemma 3 *In a Markov labour game $\mathcal{M}$, all contributing agents i must have $x_i \geq U$.*

The second result is a behaviour similar to procrastination (Rozental et al., 2022), and we have demonstrated (Theorem 1) that it is due to a combination of rationality and the opaqueness of the participating agents. Thus we have an important result: **agents will invariably take the whole time limit to complete a labour game, whether it ends in success or failure, even if $T \gg U$.**

Corollary 1 *All won Markov labour games $\mathcal{M}$ are won at time $t_{rem} = 0$.*

Connecting this result back to practical applications, if multiple agents are assigned to a burdensome task without a hard deadline, the agents will infinitely procrastinate, resulting in the task never being accomplished. Further, if the agents are required to cooperate for the completion to be preferable over failure, then the task will also never be accomplished. In either case, any agent that requires help to complete the task, due to time limits or incentive structure, will never participate. This poses a fundamental challenge for multi-agent system implementation and design, as agents will not be able to complete tasks greater than their own capabilities without an explicit third-party controller.

4.4 Mitigating Rational Procrastination

The clear obstacle for cooperation is the influence of rational procrastination preventing the agents from communicating and affecting the decision processes of others. Rational agents assume that the communication action will be wasted if cooperation is strictly required, i.e. that the other agents are minimally useful. A possible solution to this is to flip this viewpoint, and view other agents as maximally useful, or, an *optimistic* agent.

We use *maximally useful* to denote an assumption that the next t_{k_i} is favourable for as many uncommitted agents as necessary to make the thresholds t_{k_i} favourable overall (from Theorem 2). This behaviour utilizes the upper bound on favourable timesteps by using the previous observations of noncommittal action rather than the lower bound. Therefore, an optimistic agent i will 'waste' action at least as much as a non-optimistic agent, but is capable of cooperation when $x_i < U$.

Theorem 3 *The behaviour of an optimistic agent i in a Markov labour game $\mathcal{M}$ is governed by possibility of completion; an optimistic agent considers all possible commitment points, and commits if it is possible there are enough agents to critically complete the game:*

1. $T_k = \{(\lceil \frac{U-\Sigma_j u_j}{N} \rceil, N-1), (\lceil (\frac{U-\Sigma_j u_j}{N-1}) \rceil, N-2), ..., (\lceil (\frac{U-\Sigma_j u_j}{1}) \rceil, 0)\}$
2. *If* $\exists (t_{k'}, k') \in T_k \mid (t_{rem} \leq t_{k'}) \wedge (u_i + t_{rem} r_i < f_i) \wedge (t_{rem} k' < U - \Sigma_j u_j) \wedge (U - \Sigma_j u_j \leq t_{rem}(k'+1))$ *at any point, then agent i will commit*
3. *Agent i will labour iff agent i is committed and* $((t_{rem} < t_{k'}) \wedge (k_i \geq k')) \vee (t_{rem} = t_{k'})$

Proof Sketch: The proof follows similar logic to the proof of Theorem 2, with the modification for the definition of optimism represented as assuming a fictitious k' and replacing this with the observed k_i immediately after the critical timestep $t_{k'}$.

Corollary 2 *Any Markov labour game $\mathcal{M}$ with a subset of agents $\mathcal{B}$ such that $\forall i \in \mathcal{B}, |r_i| \leq |\frac{U}{|\mathcal{B}|}|$ is won by a system of optimistic agents.*

Proof Sketch: The condition states that there exists a subset of $|\mathcal{B}|$ agents for which $\frac{U}{|\mathcal{B}|}$ labour actions are favourable. This threshold will be favourable for all agents in $\mathcal{B}$ by Theorem 3, resulting in a critical completion of the game.

Optimism is a way of encoding the assumption of maximal usefulness onto the multi-agent system. In effect, this

achieves a weak form of transparency by allowing the agents to rationally cooperate when $x_i < U$. This is equivalent to introducing a mild incentive to discontinuous units of labour for each agent, $u_i = 0 \implies r_i < -\frac{U}{N}$, which is shown in the supplementary material. This shows that facilitating communication in labour games directly results in an increase in solution power for a multi-agent system, consistent with the experimental results in the literature (Tan, 1993).

5 Experiments

We illustrate several of our claims experimentally using various multi-agent reinforcement learning systems, representing our self-interested rational agents as reward-driven actors. We use standard multi-agent reinforcement learning methods, following previous literature (Schulman et al., 2017; Yu et al., 2022), except the agents do not share weights. The full details can be found in the supplementary material.

In particular, we run experiments for three cases: (1) the base case, (2) the optimistic mechanism, both described above in the analysis section; (3) the transparent case, which adds each agents' labour capacity to the observations of each agent. We use these cases to evaluate the robustness of our findings with regards to the expected labour patterns of the agents. For consistency, the heterogeneous agent labour capacities always follow $x_0 = x_1 > U > x_2 > x_3 > 1$. In order to promote learning, we introduce small stochasticity in U and T while enforcing $T > U$. We also fix all failure penalties $f_i = -1,\ \forall i \in 0, ..., N-1$ so that $H_{\text{max}} = 0$, and $H_{\text{avg}} < -1$ indicates a loss in that episode.

5.1 Results

Base Case: as shown in Fig. 1, Agents 0 and 1 labour approximately half of the labour requirement each, as expected from Theorem 2. We can also see Agents 2 and 3 labour minimally, which illustrates Lemma 3. The time remaining graph shows Theorem 1 and Lemma 1: the agents quickly converge critical completion. The average utility graph shows the favourability of the displayed labour pattern over failure. Extended plots can be found in the supplementary material.

Optimistic Mechanism Case: we have chosen to implement the optimistic mechanism as a 90% discount on the first labour action for each agent. As shown in Fig. 2, this causes the agents with smaller x_i to participate more compared to the base case. The labour penalty remains significant (e.g. Agent 3 has a strictly lower utility compared to the base case), but the first commitment threshold time is viable, therefore labour occurs temporarily as is predicted by Theorem 3.

Transparent Case: the plots for the transparent case can be found in the supplementary material. Notably, the performance is similar to the optimistic case, showing the efficiency of the mechanistic implementation.

6 Concluding Remarks & Limitations

In this work, we analyzed labour games, a class of unmediated mixed-motive cooperation games where rational agents must collectively complete a task by taking negative-reward labour actions. By studying single agents in a system of opaque agents, we find a tendency towards critical completion, a behaviour similar to procrastination. We show how this manifests in multi-agent systems, preventing contributions from less capable agents. We further illustrate the link between this behaviour and the emergence of costly communication. With these principles as a guide, we propose a mechanistic mitigation to these phenomena and analyze its effects on the multi-agent system. We demonstrate our results by encoding the problem into a multi-agent reinforcement learning system, illustrating the proposed phenomena. Finally, we raise important issues about the inefficacy of current reward structures, inviting questions on how to properly design productive multi-agent systems.

This work is not without its limitations. An envy-free solution to a labour game G can only occur in the symmetric capacity case where all agents i have $x_i \geq U$. Non-symmetric cases raise questions of fairness found in other forms of chore allocation literature (e.g. (Aziz et al., 2023; Li et al., 2023)). We leave further investigation into notions of fairness in a labour game for future works.

We also recognize that there are many equivalent formulations, including rewarding completion and rewarding idle action, which follow from an application of reward shaping (Ng et al., 1999). With p_i as the reward for taking the idle action, we propose that the following condition is sufficient for the results of our analysis to hold: $f_i < r_i < 0 \leq p_i$, and provide a proof in the supplementary material. We leave a more thorough analysis of sufficiency and necessity for future works.

The labour game we introduced in this paper is the most distilled form, intended to represent the behaviours of multi-agent systems performing division of labour on a burdensome task while accounting for the possibility of failure. This formulation is necessarily limited to a single type of task, as we do not consider aspects such as preferences between tasks with the same time thresholds or the behaviours that result. Extending the labour game with several types of tasks with separate penalties per task per agent would further increase the realism and provide more concrete insights into the best practices for multi-agent system implementation.

We also do not consider repeated labour games as a sce-

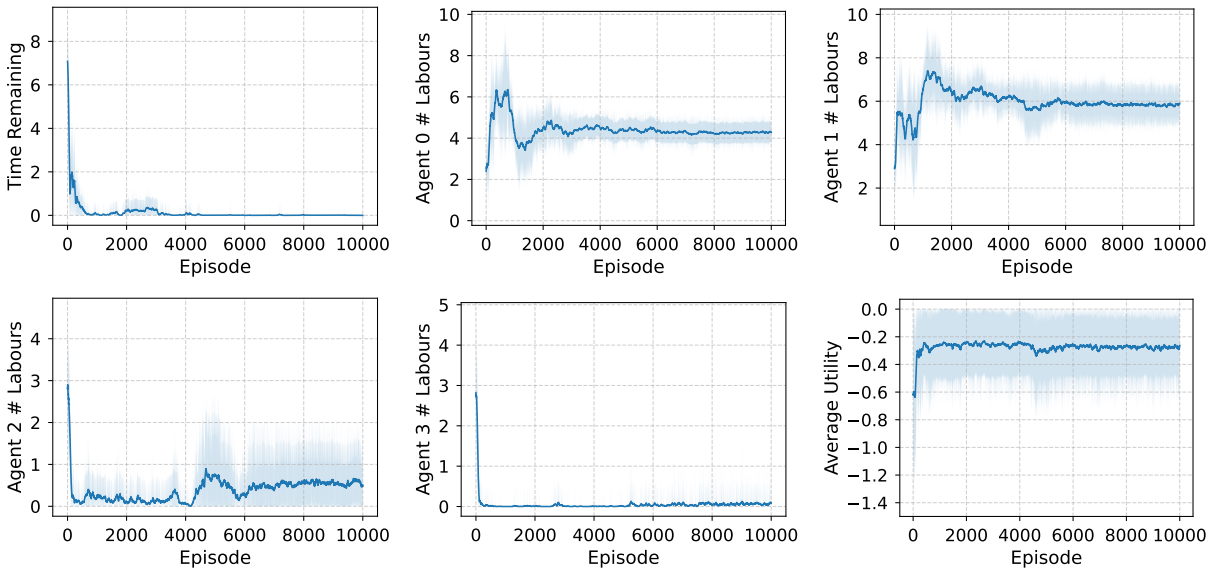

Figure 1: Agent behaviours in a Markov labour game, showing the number of times each agent took the labour action in an episode (1 game). Average utility is the average of all the agents' utility for that episode. The Time Remaining plot shows the agents quickly converge to critical completion, while the Average Utility plot shows the game is almost always won. Agents 2 and 3 are shown to contribute minimally despite a nonzero labour capacity.

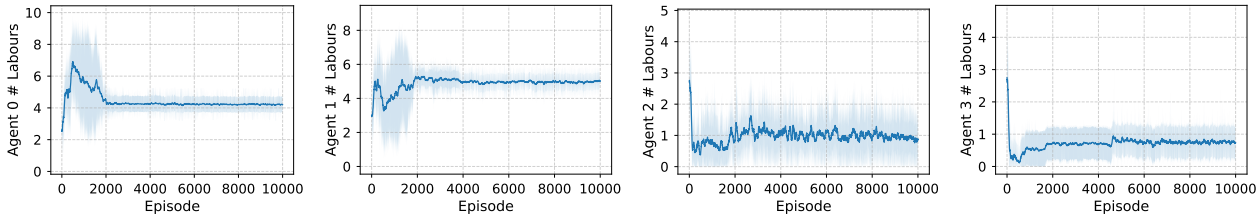

Figure 2: Agent behaviours in a Markov labour game with the optimistic mechanism (implemented as a first action discount), showing the number of times each agent took the labour action in an episode (1 game). Agents 2 and 3 are shown to contribute more despite no incentive.

nario, this would allow for more capable agents to model each other, and for strategic action to avoid taking labour actions. The repeated case is measured to a degree in the multi-agent reinforcement learning experiments, which are necessarily the same game repeated to train the policies, but we did not perform an in-depth investigation into the effects, instead using them to find the natural equilibria of the labour game. Lastly, we do not consider long-term effects from labour. The agents are assumed to be static without the ability to improve at certain tasks. An implementation of the evolution of agents may offer an alternative form of incentive to mitigate the rational procrastination phenomenon.

7 Acknowledgements

Resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute.

References

Hao Guo, Weidong Li, and Bin Deng. A survey on fair allocation of chores. *Mathematics*, 11(16), 2023. ISSN 2227-7390. doi: 10.3390/math11163616. URL https://www.mdpi.com/2227-7390/11/16/3616.

Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson, and Thore Graepel. Open problems in cooperative AI, 2020. URL https://arxiv.org/abs/2012.08630.

Jesse Clifton and Maxime Riché. Towards cooperation in learning games. Working paper, 2020.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative, multi-agent games, 2022. URL https://arxiv.org/abs/2103.01955.

Page, Robert E. and Mitchell, Sandra D. Self-organization and the evolution of division of labor. *Apidologie*, 29(1-2):171–190, 1998. doi: 10.1051/apido:19980110. URL https://doi.org/10.1051/apido:19980110.

Siyuan Liu and Jianlei Zhang. The networked division of labor game based on adaptive dynamics **this work was supported by national natural science foundation of china (grants nos. 61603201, 61603199 and 91848203), and the tianjin natural science foundation of china (grant no. 18jcy-bjc18600). *IFAC-PapersOnLine*, 52(3):156–161, 2019. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2019.06.027. URL https://www.sciencedirect.com/science/article/pii/S2405896319301119. 15th IFAC Symposium on Large Scale Complex Systems LSS 2019.

Anatol Rapoport and Albert M. Chammah. The game of chicken. *American Behavioral Scientist*, 10(3):10–28, 1966. doi: 10.1177/000276426601000303. URL https://doi.org/10.1177/000276426601000303.

Raphael Memmesheimer, Martina Overbeck, Bjoern Kral, Lea Steffen, Sven Behnke, Martin Gersch, and Arne Roennau. Cleaning robots in public spaces: A survey and proposal for benchmarking based on stakeholders interviews, 2024. URL https://arxiv.org/abs/2407.16393.

Gabriel Dulac-Arnold, Daniel J. Mankowitz, and Todd Hester. Challenges of real-world reinforcement learning. *CoRR*, abs/1904.12901, 2019. URL http://arxiv.org/abs/1904.12901.

John Nash. Non-cooperative games. *Annals of Mathematics*, 54(2):286–295, 1951. ISSN 0003486X, 19398980. URL http://www.jstor.org/stable/1969529.

Darius Muglich, Luisa Zintgraf, Christian Schroeder de Witt, Shimon Whiteson, and Jakob Foerster. Generalized beliefs for cooperative AI, 2022. URL https://arxiv.org/abs/2206.12765.

Bowen Baker. Emergent reciprocity and team formation from randomized uncertain social preferences, 2020. URL https://arxiv.org/abs/2011.05373.

Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. Cooperative ai: machines must learn to find common ground. *Nature*, 593(7857):33–36, 2021.

Ming Tan. Multi-agent reinforcement learning: Independent versus cooperative agents. In Paul E. Utgoff, editor, *Machine Learning, Proceedings of the Tenth International Conference, University of Massachusetts, Amherst, MA, USA, June 27-29, 1993*, pages 330–337. Morgan Kaufmann, 1993. doi: 10.1016/B978-1-55860-307-3.50049-6. URL https://doi.org/10.1016/b978-1-55860-307-3.50049-6.

Richard Willis, Yali Du, Joel Z Leibo, and Michael Luck. Will systems of llm agents cooperate: An investigation into a social dilemma, 2025. URL https://arxiv.org/abs/2501.16173.

Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, Heather Roff, and Thore Graepel. Inequity aversion improves cooperation in intertemporal social dilemmas. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/7fea637fd6d02b8f0adf6f7dc36aed93-Paper.pdf.

Kunlun Zhu, Hongyi Du, Zhaochen Hong, Xiaocheng Yang, Shuyi Guo, Zhe Wang, Zhenhailong Wang, Cheng Qian, Xiangru Tang, Heng Ji, and Jiaxuan You. Multiagentbench: Evaluating the collaboration and competition of llm agents, 2025. URL https://arxiv.org/abs/2503.01935.

Lewis Hammond, Alan Chan, Jesse Clifton, Jason Hoelscher-Obermaier, Akbir Khan, Euan McLean, Chandler Smith, Wolfram Barfuss, Jakob Foerster, Tomáš Gavenčiak, The Anh Han, Edward Hughes, Vojtěch Kovařík, Jan Kulveit, Joel Z. Leibo, Caspar Oesterheld, Christian Schroeder de Witt, Nisarg Shah, Michael Wellman, Paolo Bova, Theodor Cimpeanu, Carson Ezell, Quentin Feuillade-Montixi, Matija Franklin, Esben Kran, Igor Krawczuk, Max Lamparth, Niklas Lauffer, Alexander Meinke, Sumeet Motwani, Anka Reuel, Vincent Conitzer, Michael Dennis, Iason Gabriel, Adam Gleave, Gillian Hadfield, Nika Haghtalab, Atoosa Kasirzadeh, Sébastien Krier, Kate Larson, Joel Lehman, David C. Parkes, Georgios Piliouras, and Iyad Rahwan. Multi-agent risks from advanced AI, 2025. URL https://arxiv.org/abs/2502.14143.

Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrinmaya Sachan, and Rada Mihal-

cea. Cooperate or collapse: Emergence of sustainable cooperation in a society of llm agents, 2024. URL https://arxiv.org/abs/2404.16698.

Vincent Conitzer and Caspar Oesterheld. Foundations of cooperative AI. In Brian Williams, Yiling Chen, and Jennifer Neville, editors, *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 15359–15367. AAAI Press, 2023. doi: 10.1609/AAAI.V37I13.26791. URL https://doi.org/10.1609/aaai.v37i13.26791.

Shanshan Han, Qifan Zhang, Yuhang Yao, Weizhao Jin, Zhaozhuo Xu, and Chaoyang He. Llm multi-agent systems: Challenges and open problems, 2024. URL https://arxiv.org/abs/2402.03578.

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, Xiaojun Chang, Junge Zhang, Feng Yin, Yitao Liang, and Yaodong Yang. Proagent: Building proactive cooperative agents with large language models, 2024. URL https://arxiv.org/abs/2308.11339.

Arrasy Rahman, Ignacio Carlucho, Niklas Höpner, and Stefano V. Albrecht. A general learning framework for open ad hoc teamwork using graph-based policy learning, 2023. URL https://arxiv.org/abs/2210.05448.

Jianhong Wang, Yang Li, Yuan Zhang, Wei Pan, and Samuel Kaski. Open ad hoc teamwork with cooperative game theory, 2024. URL https://arxiv.org/abs/2402.15259.

Xudong Guo, Kaixuan Huang, Jiale Liu, Wenhui Fan, Natalia Vélez, Qingyun Wu, Huazheng Wang, Thomas L. Griffiths, and Mengdi Wang. Embodied llm agents learn to cooperate in organized teams, 2024. URL https://arxiv.org/abs/2403.12482.

Brian Mintz and Feng Fu. Evolutionary multi-agent reinforcement learning in group social dilemmas, 2024. URL https://arxiv.org/abs/2411.10459.

Woodrow Z. Wang, Mark Beliaev, Erdem Bıyık, Daniel A. Lazar, Ramtin Pedarsani, and Dorsa Sadigh. Emergent prosociality in multi-agent games through gifting, 2021. URL https://arxiv.org/abs/2105.06593.

Edward Hughes, Thomas W. Anthony, Tom Eccles, Joel Z. Leibo, David Balduzzi, and Yoram Bachrach. Learning to resolve alliance dilemmas in many-player zero-sum games, 2020. URL https://arxiv.org/abs/2003.00799.

Julian Yocum, Phillip Christoffersen, Mehul Damani, Justin Svegliato, Dylan Hadfield-Menell, and Stuart Russell. Mitigating generative agent social dilemmas. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*, 2023. URL https://openreview.net/forum?id=5TIdOk7XQ6.

Andrei Lupu and Doina Precup. Gifting in multi-agent reinforcement learning. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, page 789–797, Richland, SC, 2020. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450375184.

Manuel Rios, Nicanor Quijano, and Luis Felipe Giraldo. Understanding the world to solve social dilemmas using multi-agent reinforcement learning, 2023. URL https://arxiv.org/abs/2305.11358.

Joel Z. Leibo, Edgar A. Duéñez-Guzmán, Alexander Vezhnevets, John P. Agapiou, Peter Sunehag, Raphael Koster, Jayd Matyas, Charlie Beattie, Igor Mordatch, and Thore Graepel. Scalable evaluation of multi-agent reinforcement learning with melting pot. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 6187–6199. PMLR, 2021. URL http://proceedings.mlr.press/v139/leibo21a.html.

Elizaveta Tennant, Stephen Hailes, and Mirco Musolesi. Modeling moral choices in social dilemmas with multi-agent reinforcement learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI-2023, page 317–325. International Joint Conferences on Artificial Intelligence Organization, August 2023. doi: 10.24963/ijcai.2023/36. URL http://dx.doi.org/10.24963/ijcai.2023/36.

Kevin R. McKee, Edward Hughes, Tina O. Zhu, Martin J. Chadwick, Raphael Koster, Antonio Garcia Castaneda, Charlie Beattie, Thore Graepel, Matt Botvinick, and Joel Z. Leibo. A multi-agent reinforcement learning model of reputation and cooperation in human groups, 2023. URL https://arxiv.org/abs/2103.04982.

Andreas A. Haupt, Phillip J. K. Christoffersen, Mehul Damani, and Dylan Hadfield-Menell. Formal contracts mitigate social dilemmas in multi-agent rl, 2024. URL https://arxiv.org/abs/2208.10469.

Justin Chih-Yao Chen, Swarnadeep Saha, and Mohit Bansal. Reconcile: Round-table conference improves reasoning via consensus among diverse llms, 2024. URL https://arxiv.org/abs/2309.13007.

Bo Li, Fangxiao Wang, and Yu Zhou. Fair allocation of indivisible chores: Beyond additive costs, 2023. URL https://arxiv.org/abs/2205.10520.

Jugal Garg, Aniket Murhekar, and John Qin. New algorithms for the fair and efficient allocation of indivisible chores. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI-2023, page 2710–2718. International Joint Conferences on Artificial Intelligence Organization, August 2023. doi: 10.24963/ijcai.2023/302. URL http://dx.doi.org/10.24963/ijcai.2023/302.

Benjamin Cookson, Soroush Ebadian, and Nisarg Shah. Constrained fair and efficient allocations, 2024. URL https://arxiv.org/abs/2411.00133.

Lang Yin and Ruta Mehta. On the envy-free allocation of chores, 2022. URL https://arxiv.org/abs/2211.15836.

Haris Aziz, Jeremy Lindsay, Angus Ritossa, and Mashbat Suzuki. Fair allocation of two types of chores, 2023. URL https://arxiv.org/abs/2211.00879.

Umang Bhaskar, A. R. Sricharan, and Rohit Vaish. On approximate envy-freeness for indivisible chores and mixed resources, 2022. URL https://arxiv.org/abs/2012.06788.

Soroush Ebadian, Rupert Freeman, and Nisarg Shah. Efficient resource allocation with secretive agents. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 272–278. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/39. URL https://doi.org/10.24963/ijcai.2022/39. Main Track.

Zhengwu Zhao and Chunyan Zhang. The mechanisms of labor division from the perspective of task urgency and game theory. *Physica A: Statistical Mechanics and its Applications*, 630:129284, 2023. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2023.129284. URL https://www.sciencedirect.com/science/article/pii/S0378437123008397.

Alexander Rozental, David Forsström, Ayah Hussoon, and Katrin B. Klingsieck. Procrastination among university students: Differentiating severe cases in need of support from less severe cases. *Frontiers in Psychology*, Volume 13 - 2022, 2022. ISSN 1664-1078. doi: 10.3389/fpsyg.2022.783570. URL https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.783570.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. URL https://arxiv.org/abs/1707.06347.

Andrew Y Ng, Daishi Harada, and Stuart J Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 278–287, 1999.

It's Rational for AI Agents to Procrastinate

Supplementary Material

Contents

This supplementary material is composed of three sections. In Section A, we provide restatements and full proofs for all lemmas, theorems, corollaries, and propositions from the main body of the paper. In Section B, we provide the training details, including compute, hyperparameters, and the encoding of the Markov labour game into a multi-agent reinforcement learning environment. In Section C, we provide the extended plots from the three cases analyzed, along with an explanation of the behaviour represented by the plots.

A Proofs

A.1 Proposition 1

Statement: An agent i in a Markov labour game $\mathcal{M}$ will win the game if and only if $x_i \geq U$, and will otherwise idle rather than labouring.

Proof: The utility of an agent i if the game is lost is: $H_{i_{\text{loss}}} = r_i u_i + f_i$, and if the game is won: $H_{i_{\text{win}}} = r_i U$. As we have $f_i < r_i < 0$ from the assumptions (and $u_i \geq 0$), both terms in $H_{i_{\text{loss}}}$ are negative. The agent i operates on the principle of utility maximization, so in the case of a loss we have $u_i = 0$. Therefore, the utility of a single agent i is $H_i = \max(r_i U,\ f_i)$, with the first case representing a win and the second case representing a loss. The requirement for a win is then $r_i U \geq f_i$ or $U \leq \frac{f_i}{r_i} = x_i$. ■

A.2 Lemma 1

Statement: In a Markov labour game $\mathcal{M}$, an agent i acting in isolation will have at most one action discontinuity.

Proof: This behaviour is consistent with the behaviour of delaying action. The number of timesteps in the game where outside action could possibly occur (thus minimizing u_i for agent i) is maximized. However, there is no new information gathered by agent i at any point during the length of the game, so the timing of action discontinuities is equally favourable, with subsequent action discontinuities serving no purpose. Combining these two facts, the action discontinuity will occur only once if $x_i \geq U$ and zero times otherwise. ■

A.3 Theorem 1

Statement: An agent i acting in a Markov labour game $\mathcal{M}$ without knowledge of other agents' incentive structures will critically complete all labour games where $x_i \geq U$, where x_i is the labour capacity defined Definition 4.

Proof: Following from Lemma 1 and Prop 1, it is preferable for any given agent to act later than earlier to minimize their amount of labour due to the possibility of intervention. A single agent will therefore idle for the first $T - U$ timesteps, at which point another idle action directly results in failure. Thus labour will always begin at $t_{rem} = U$ in the absence of interference. If any labour is performed by an outside source before $t_{rem} = U$, this forms another game with $U' = U - \Sigma_j u_j$ and a new action discontinuity point of $t_{rem} = U'$. If labour is performed by an outside source after $t_{rem} = U$, then it follows from the same logic that any further labour from agent i is pushed back the same amount. This argument can be continued successively until the last timestep. The last unit of labour is performed on the last timestep in all cases, verifying the claim. ■

A.4 Lemma 2

Statement: An agent i acting in multi-agent system inside a Markov labour game $\mathcal{M}$ will minimize action discontinuity.

Proof: From Theorem 1, the agent will critically complete the game and maximize idle actions in the process. It is guaranteed then that a unit of labour will occur on the last actionable timestep $t_{rem} = 1$; consider a sub-game with $U' = U - 1$ and $T' = T - 1$. Using an inductive argument, the agent will critically complete this new game by performing labour on the last actionable timestep ($t_{rem} = 2$ in the original game). This argument can be extended until $t_{rem} = U$, which accounts for all units of labour. In this case, there is only one action discontinuity. If labour is performed by agents other than agent i before $t_{rem} = U$, then agent i will still only have one action discontinuity, as the point of discontinuity is shifted back by an equal amount to retain critical completion. This is the minimum number of action discontinuities for a won game, as stated in Lemma 1. If labour is performed by agents other than agent i after $t_{rem} = U$, then one action discontinuity per contiguous interval of outside labour is necessary to retain critical completion, but no more. This is the minimum amount of action discontinuities to retain critical completion (Theorem 1) without outside knowledge of the labour patterns of outside agents. In all cases, the number of action discontinuities is minimized. ■

A.5 Theorem 2

Statement: Let k_i be the number of committed agents, excluding agent i, in the Markov labour game $\mathcal{M}$, and t_{k_i} be the commitment decision times for an agent i. The behaviour of an agent i in $\mathcal{M}$ is governed by the number of committed agents; an uncommitted agent will only commit if they are necessary and sufficient to critically complete the

game:

1. $t_{k_i} = \lceil \frac{U - \Sigma_j u_j}{k_i + 1} \rceil$
2. If $(t_{rem} \leq t_{k_i}) \wedge (u_i + t_{rem} r_i < f_i) \wedge (t_{rem} k_i < U - \Sigma_j u_j) \wedge (U - \Sigma_j u_j \leq t_{rem}(k_i + 1))$, at any point, then agent i will commit
3. Agent i will labour iff i is committed and $t_{rem} \leq t_{k_i}$

Proof: From Theorem 1, any agent i will critically complete a Markov labour game $\mathcal{M}$ if $x_i \geq U$. This behaviour accounts for the labour of other agents through a modified commitment point, which is expanded upon in Lemma 2. The mathematical representation of the commitment point for an agent i relies upon the promise of future labour by other agents (their commitment) and the total work remaining. The ratio of these values gives the amount of labour that each of the committed agents (other than agent i) plus agent i would have to take on to critically complete the game. This is captured in the first statement. Further, agents will maximize their overall reward by minimizing the number of labour actions taken. The second statement states a committing agent must: (1) be in critical completion range (Theorem 3), (2) have completion be preferable based on future and past labour, (3) be sufficient for the critical completion of the game, and (4) be necessary for the critical completion of the game. This is consistent with the results seen so far. The third statement governs the timing of the actual labour actions once committed. Commitment (from Definition 6) states that labouring contiguously for the remaining timesteps is preferable for agent i if agent i is committed; however, this is not optimal when considering $k_i > 0$. At each timestep where labour is performed, t_{k_i} updates so that critical completion is preserved, leading directly to statement 3. ■

A.6 Lemma 3

Statement: In a Markov labour game $\mathcal{M}$, all contributing agents i must have $x_i \geq U$.

Proof: The proof directly follows from Theorem 2; agents will never commit after t_0, as each agent that labours at t_0 can critically complete the game. All agents that do not labour at t_0 will never be necessary to critically complete the game and will not commit. Therefore, all agents that will take the labour action at least once, will take a labour action at t_0, and so must have $x_i \geq U$. ■

A.7 Corollary 1

Statement: All won Markov labour games $\mathcal{M}$ are won at time $t_{rem} = 0$.

Proof: The proof follows directly from Theorem 1 and Theorem 2. Critical completion is always preserved through the rules governing the labour patterns of the agents. Therefore, the game will always end at $t_{rem} = 0$, and all won Markov labour games are won at this time as well. ■

A.8 Theorem 3

Statement: The behaviour of an optimistic agent i in a Markov labour game $\mathcal{M}$ is governed by possibility of completion; an optimistic agent considers all possible commitment points, and commits if it is possible there are enough agents to critically complete the game:

1. $T_k = \{(\lceil \frac{U - \Sigma_j u_j}{N} \rceil, N - 1), (\lceil (\frac{U - \Sigma_j u_j}{N-1}) \rceil, N - 2), ..., (\lceil (\frac{U - \Sigma_j u_j}{1}) \rceil, 0)\}$
2. If $\exists (t_{k'}, k') \in T_k \mid (t_{rem} \leq t_{k'}) \wedge (u_i + t_{rem} r_i < f_i) \wedge (t_{rem} k' < U - \Sigma_j u_j) \wedge (U - \Sigma_j u_j \leq t_{rem}(k' + 1))$ at any point, then agent i will commit
3. Agent i will labour iff agent i is committed and $((t_{rem} < t_{k'}) \wedge (k_i \geq k')) \vee (t_{rem} = t_{k'})$

Proof: The proof follows similar logic to the proof of Theorem 2, with the modification for the definition of optimism. Optimistic agents require only one of these timesteps to be possibly valid for commitment after the relevant timestep, accounting for the possibility that other agents may commit at that timestep (as opposed to the current t_{k_i} needing to be valid already in Theorem 2). This is equivalent to assuming a fictitious k' and replacing this with the observed k_i immediately after the critical timestep $t_{k'}$. The agent-specific timestep t_{k_i} is replaced with the set of all possible (t_k, k) pairs, represented as T_k. The agent then only labours if the fictitious k' matches the observed k_i after the timestep. ■

A.9 Corollary 2

Statement: Any Markov labour game $\mathcal{M}$ with a subset of agents $\mathcal{B}$ such that $\forall i \in \mathcal{B}, |r_i| \leq |\frac{U}{|\mathcal{B}|}|$ is won by a system of optimistic agents.

Proof: The condition states that there exists a subset of $|\mathcal{B}|$ agents for which $\frac{U}{|\mathcal{B}|}$ labour actions are favourable. This threshold will be favourable for all agents in $\mathcal{B}$ by Theorem 3, resulting in a critical completion of the game. ■

A.10 Sufficient Condition for Results to Hold

Statement: With p_i as the reward for taking the idle action, we propose that the following condition is sufficient for the results of our analysis to hold: $f_i < r_i < 0 \leq p_i$

Explanation: The conditions $f_i < r_i < 0$ are assumed in the analysis of this problem. The remaining conditions to preserve are (1) the maximization of idle actions and (2) the preference of winning the game over failure in the

presence of these modified idle rewards. The condition $0 \leq p_i$ ensures that the maximization of idle actions (as $f_i, r_i < 0$). The condition from the original formulation ensuring the second property is $f_i < r_i$, so that some amount of labour is preferred over failure. The general form of this inequality is $f_i < r_i - p_i$ as taking the labour action now has an element of lost utility from not taking the idle action. With properties (1) and (2) preserved, all of the behaviour from the base game is also preserved. ■

B Experimental Procedures

B.1 Training Details

Hyperparameter	Value
anneal_lr	True
batch_size	128
clip_coef	0.2
clip_vloss	True
ent_coef	0.01
gae_lambda	0.95
gamma	1
learning_rate	0.00025
max_grad_norm	0.5
minibatch_size	32
n_agents	2
norm_adv	True
num_iterations	1562
num_minibatches	4
num_steps	128
target_kl	None
total_timesteps	200000
update_epochs	4
vf_coef	0.5

Table 1: Reinforcement learning hyperparameters

We used one RTX 6000, 4 CPU cores, and 20 GB of CPU RAM to run these experiments. The total time elapsed for four runs was approximately three hours. All runs used a standard implementation of proximal-policy-optimization with the same hyperparameters (given in Table 1) with separate weights and policies per agent.

B.2 Experimental Details

We encode the Markov labour game $\mathcal{M}$ into a multi-agent reinforcement learning environment using PettingZoo. The observations for an agent i in the base case is $(r_i, U - \Sigma_j u_j, t_{rem}, \mathbf{a}, k_i)$ where $\mathbf{a}$ is the ordered action tuple from the previous timestep, describing the actions taken by each agent in the previous timestep. The optimistic mechanism case uses the same observation space as the base case. The transparency case adds the labour penalties of each other agent to the observations, $(\mathbf{r}, U - \Sigma_j u_j, t_{rem}, \mathbf{a}, k_i)$ where $\mathbf{r}$ is the ordered labour penalty tuple that gives each agent's labour penalty.

All experiments were run with a constant penalty r_i for each agent i. We give U (the labour required) some stochasticity to promote learning and discourage memorization: $U = U_{base} + \delta_U$ where δ_U is a randomly sampled integer $\delta_U \in \{0, 1, 2\}$ and $U_{base} = 10$. The labour penalties are $r_0 = r_1 = -1/12$, $r_2 = -1/6$, $r_3 = -1/2$. These parameters were chosen in relation to U and f_i (which are set as $f_i = -1$ as described in the Experiments section). Agents 0 and 1 are specified such that they have $x_0, x_1 \geq U$ and agents 2 and 3 are varying levels of less capable. The time limit T is set based on the end value of U, specifically $T = U + \max(\delta_T) + \delta_T$, where δ_T is a randomly sampled integer $\delta_T \in \{0, 1, 2\}$.

Each plot (in the body and in the supplementary material) is taken as an average of 4 runs with 1-sigma error bars shown. All runs used 200 000 timesteps (approximately 10 000 episodes) and were verified for convergence.

C Plots

In this section, we provide supplementary plots fully detailing the behaviours of each agent in the Markov labour game. The number of labour actions per episode plots are repeated from the body of the paper for completeness.

C.1 Base Case

The base case has no modifications, it is a direct implementation of the Markov labour game $\mathcal{M}$. In Figure 3, we can see several of the behaviours from the analysis section illustrated. The "Time Remaining" plot shows Corollary 1. The "Agent i Labours" graphs show that the major contributors (agents 0 and 1) have $x_i \geq U$ (Theorem 2), while the remaining agents ($x_i < U$) take the labour action less than once per episode on average. The agent "Utility" plots show how agents 2 and 3 exploit agents 0 and 1 by utilizing less of their labour capacity. Agents 0 and 1 share a similar amount of the penalty while agents 2 and 3 face almost no penalties. The average utilities do not reach $f_i = -1$ within error, indicating that the game is rarely lost.

C.2 Optimistic Mechanism Case

The optimistic mechanism case is implemented as a 90% discount on the first labour penalty, allowing each agent to act as if they have $x'_i = 10x_i$ for this first action. We retain the reward structure from the base case ($x_0 = x_1 \geq U > x_2 > x_3$), but we see in Figure 4 that agents 2 and 3 contribute meaningfully (around 1 labour action per episode) despite their labour capacities (Theorem 3). The utility plots

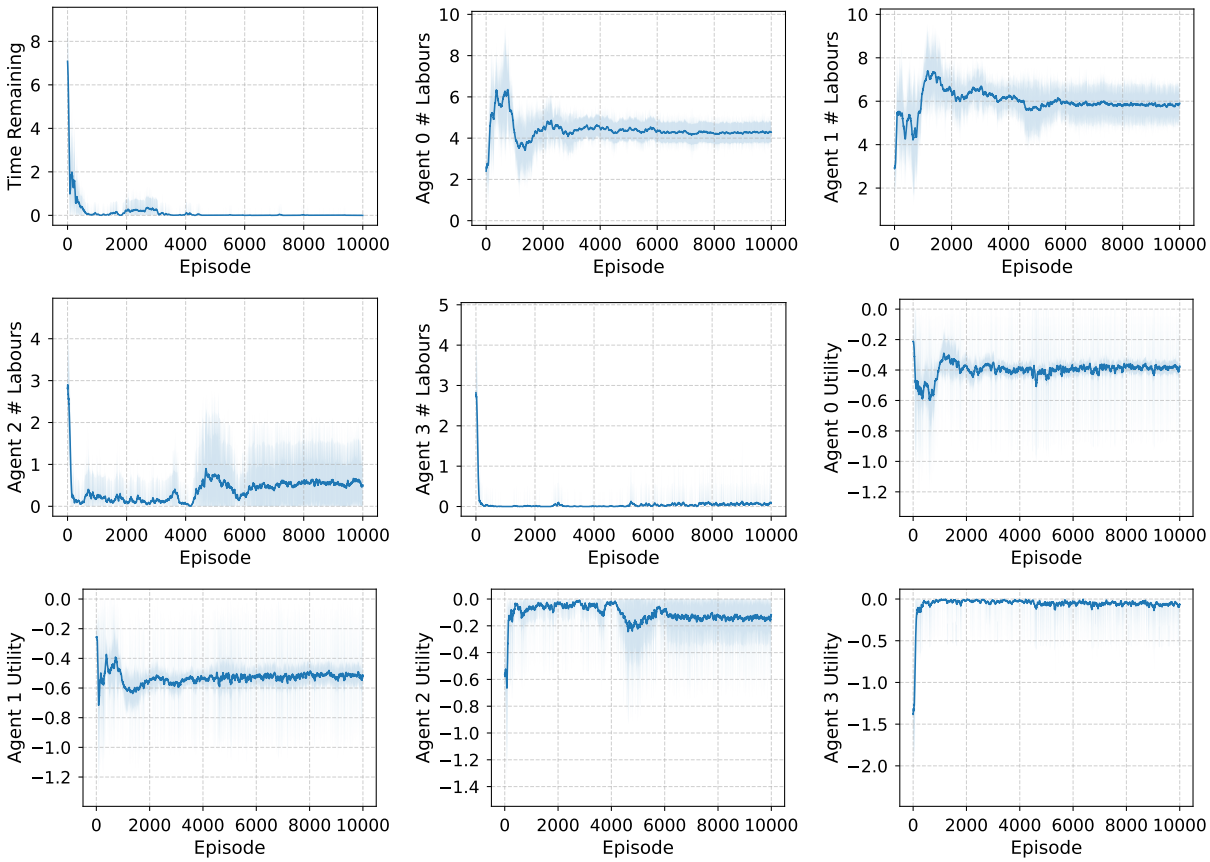

Figure 3: Agent statistics in a Markov labour game without modifications, showing the number of times each agent took the labour action in an episode (1 game). The Time Remaining plot shows the agents quickly converge to critical completion. Agents 2 and 3 are shown to contribute minimally despite the capacity to do so, and have higher utilities than agents 0 and 1.

show that this is closer to an equality-based solution as the average utility is less negative and the variance between the agents is lower. The time remaining graph shows that the agents still have a preference towards critical completion (Theorem 1).

C.3 Transparent Case

The transparent case adds the labour penalties of each agent to every agent's observation space, giving each agent full knowledge of $x_0, ..., x_{N-1}$. We retain the reward structure from the base case ($x_0 = x_1 \geq U > x_2 > x_3$), but we see in Figure 5 that agents 2 and 3 contribute meaningfully despite their labour capacities (Theorem 3). In particular, agent 2 is a significant contributor. The utility plots show that this is closer to an equality-based solution than the optimistic mechanism or base cases as the utilities converge to a similar value. However, the average utility appears to be lower than the base or optimistic mechanism cases, which is explained by a higher failure rate. This higher failure rate is likely due to the full transparency transforming the labour game into a game of chicken. The time remaining graph shows that the agents still have a preference towards critical completion when the game is won (Theorem 1).

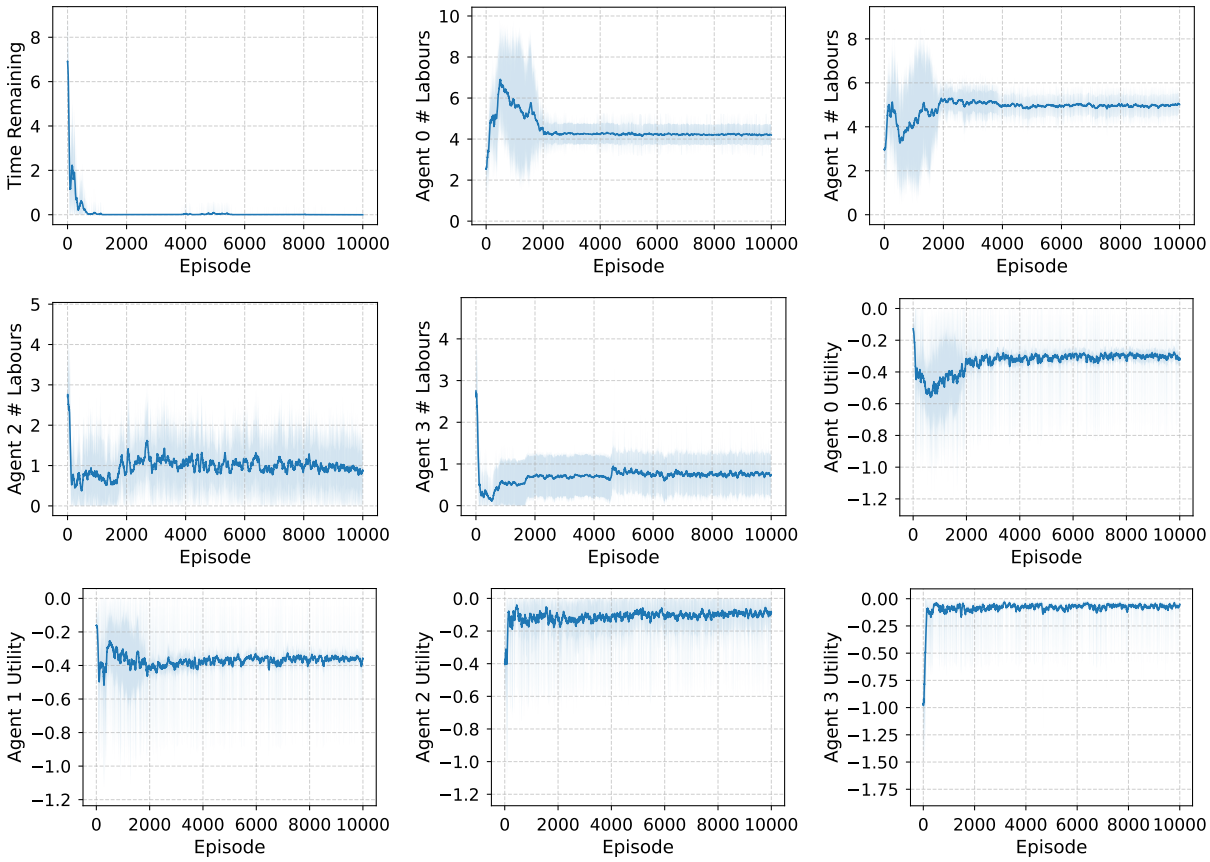

Figure 4: Agent statistics in a Markov labour game with the optimistic mechanism modification, showing the number of times each agent took the labour action in an episode (1 game). The Time Remaining plot shows the agents still quickly converge to critical completion. Compared to the base case, Agents 2 and 3 contribute significantly more, and agents 0 and 1 have a less negative utility.

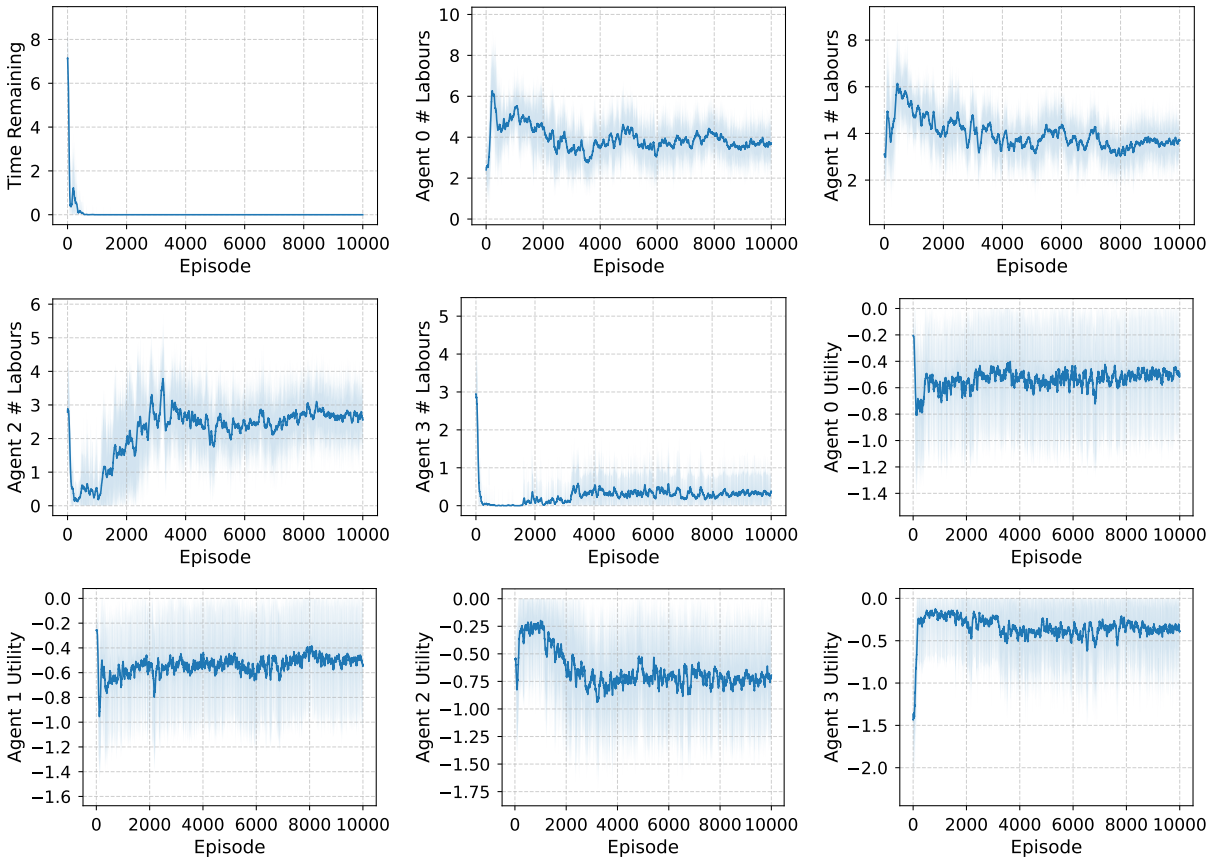

Figure 5: Agent behaviours in a Markov labour game with the transparent modification, showing the number of times each agent took the labour action in an episode (1 game). The Time Remaining plot shows the agents still quickly converge to critical completion. Compared to the base case, Agents 2 and 3 contribute significantly more. The performance is similar to the optimistic mechanism case, showing the efficacy of the mechanistic solution.