The AI economist: Improving Equality and Productivity with AI-Driven Tax Policies

Stephan Zheng, Alexander Trott, Sunil Srinivasa, Nikhil Naik, Melvin Gruesbeck, David C. Parkes, and Richard Socher, 2020, mimeo.

発表者: 冨田 燿志

2021 年 6 月 21 日 PaperFriday

1. イントロダクション

- 最適な課税・再分配制度のデザインは重要だが難しい問題である.
- その難しさの一因は, 生産性と平等性のトレードオフ:
  - 税を低くすると, 経済行動が活発になり生産性は増すが不平等になる.
  - 税を高く再分配を厚くすると, 経済行動が抑制され生産性が落ちる.

- 最適課税理論の分野で研究が進むが, 明確な答えは得られていない:
  - 税率変化の経済行動への影響（弾力性）の推定が困難.
  - 税制をフィールド実験することはほぼ不可能.

- この論文では, 経済シミュレーションゲームにおいて, 政府主体に最適な税制を深層強化学習により学習させることを考える（AI Economist）.
- 市民を RL エージェントとするシミュレーション実験と, 人を被験者とする実験の双方で, AI economist による税制は生産性・平等性のトレードオフを高い水準で達成した.

2. Gather-and-Build Games

Gather-and-Build game（収集建築ゲーム）

- 2 次元のグリッド (25×25) からなる世界が舞台.
- 市民はフィールドを歩き回り, 資源（石と木）を集め, それらを 1 つずつ使って家を建てコインを稼ぎ, またコインを介して資源の取引をする.
- 資源は空タイルに確率的に産み出される.
- 市民は家を建てるとコインが得られるが, 得られるコインの数は各市民のスキルごとに異なる.
- （政府主体・税制度は後述）

労働とスキル.

- 市民の行動（移動, 収集, 取引, 建築）にはそれぞれ労働コストが設定されている.
- 各期に市民がどれか 1 つの行動をとると, 設定されている労働コストがかかる.

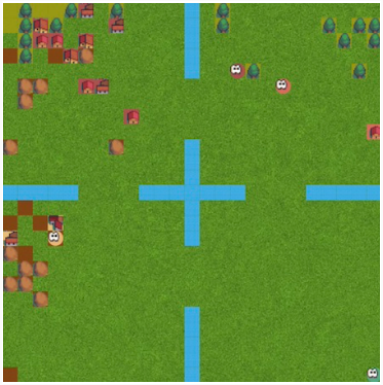
- 建築には, 木と石の資源が 1 つずつ必要.
- 建築スキル（1 以上 3 以下）が各市民に設定されていて, 建築をすると市民は $10 \times$ スキル 枚のコインを得る.

取引.

- 取引を選択した市民は, 市場に売り注文「コイン X 枚以上で木を売ります」か, 買い注文「コイン Y 枚以下で石を買います」を送る.
- 市場に売り注文買い注文が出れば, 先に注文を出した側の値段で自動的に取引がされる.
  - 複数ある場合はより条件の良い注文が優先される.
  - 一定期間取引相手が見つからなかった注文は市場から削除される.

シナリオ.

- フィールドは水により 4 つの区域に別れている（水部分は通れない）
- 資源は空間的に集まって発生する.
- 市民は 4 人.
- 建築スキルは 1.13, 1.33, 1.65, 22.2 で固定（パレート分布の分位点を元に設定）
- 1 エピソードは, 1000 ステップからなる.
- 市民は自身の位置の近く（11×11）のみの状況が観測できる.



2.1 市民の最適化行動

- 市民の効用関数:

$$u_i(x_{i,t}, l_{i,t}) = \frac{(x^c_{i,t})^{1-\eta} - 1}{1-\eta} - l_{i,t}. \tag{2}$$

  - $x_{i,t} = (x^w_{i,t}, x^s_{i,t}, x^c_{i,t})$: 市民 i が t 期時点で保有する木・石・コイン.
  - $l_{i,t}$: t 期までの蓄積労働量.
  - $\eta \in (0,1)$: 市民の効用関数の非線形性をコントロールするパラメータ.
- 合理的な市民は以下の最大化を行う.

$$\forall i : \max_{\pi_i} \mathbb{E}_{a_i \sim \pi_i,\ \boldsymbol{a}_{-i} \sim \boldsymbol{\pi}_{-i}, s' \sim \mathscr{T}} \left[u_i(x_{i,0}, l_{i,0}) + \sum_{t=1}^{H} \gamma^t \underbrace{\left(u_i(x_{i,t}, l_{i,t}) - u_i(x_{i,t-1}, l_{i,t-1})\right)}_{=r_{i,t}} \right]. \tag{3}$$

  - π_i: 状態 $s \in \mathscr{T}$ と観測 $o_{i,t}$ から行動 $a_{i,t}$ を選ぶポリシー
  - $\gamma \in (0,1)$: 割引因子

深層強化学習エージェント

- ディープニューラルネットワークを用いて市民のポリシーをモデリングする:

$$a_{i,t} \sim \pi(o_{i,t}^{\text{world}}, o_{i,t}^{\text{agent}}, o_{i,t}^{\text{market}}, o_{i,t}^{\text{tax}}, h_{i,t-1}; \theta)$$

  - $o_{i,t}^{\text{world}}$: 近くの状況に関する観測.
  - $o_{i,t}^{\text{agent}}$: 市民の状況 (資源・コイン保有)
  - $o_{i,t}^{\text{market}}$: 市場の状況（売り注文・買い注文の蓄積状況）
  - $o_{i,t}^{\text{tax}}$: 税率（後述）
  - $h_{i,t-1}$: hidden state（自身のプライベートな状況 (スキルと労働蓄積量) と過去の履歴）
  - θ: モデルのパラメータ

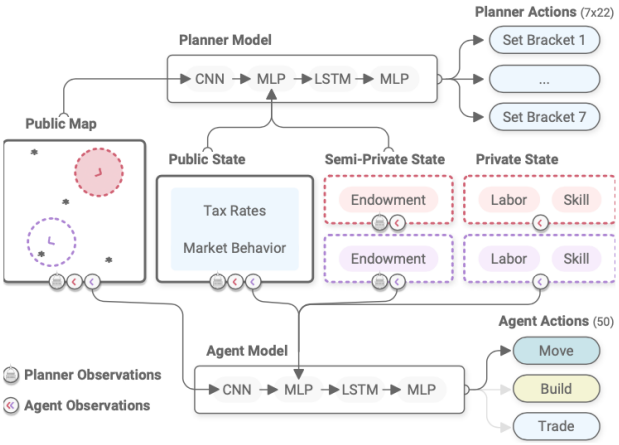
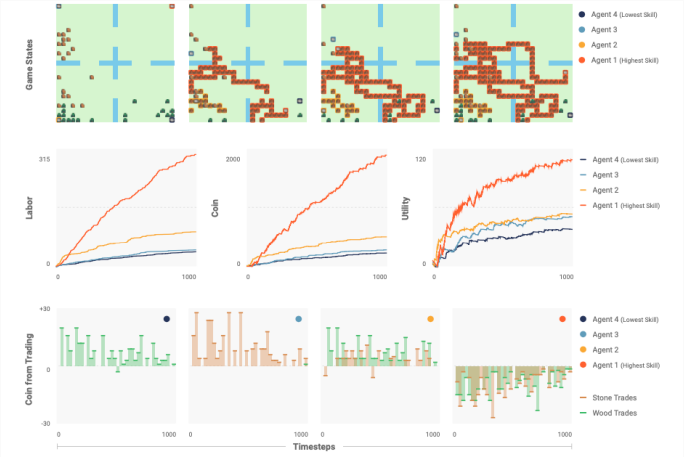


Figure 3: Schematic overview of the general network architecture used in our work. Spatial observations are processed by a stack of two convolutional layers (CNN) and flattened into a fixed-length feature vector. This feature vector is concatenated with the remaining observation inputs and the result is processed by a stack of two fully connected layers (MLP). The output is then used to update the hidden state of an LSTM and action logits are computed via a linear projection of the updated hidden state. Finally, the network computes a softmax probability layer for each action head. For the agent policy, there is a single action space and action head. For the tax policy, there is a separate action space and action head for each tax rate the tax policy controls (described below).

無課税・無分配下での市民の行動



- 左図は無課税下で学習後の各市民の１エピソードの行動.
- 低スキル市民（紺, 水）は資源を集めて市場で売ることに徹している.
- 高スキル市民（オレンジ）は市場で資源を買い, 自身で建築してコインを稼ぐ.
- 黄色は最初は建築しているが, のちに資源を売る方にスイッチしている.

3. 最適課税政策の学習

- 課税と再分配（のみ）を行う政府主体を導入する.
- 政府は, 生産性と平等性のトレードオフに直面している.
  - 無課税・無分配では生産性は最大化されるが, 不平等が生まれる.
  - 課税・再分配を行うと平等性が増すが, 生産性が落ちる.

- ここでは, 無課税, 米連邦所得税, Saez フレームワーク, AI economist の 4 つの税制度を試す.

税ピリオドと再分配.

- 税ピリオドは M ステップ続く（実験では $M = 100$ とし, 1 エピソードに 10 ピリオドある）
- ピリオド p の税は, 期初 t から 期末 $t + M$ までの収入 z_i^p に課される.

- ピリオドの初めに, 政府は税額関数 $T(z)$ を決めて公表する.
  - 各市民 i は, 収入 z_i^p に応じて $T(z_i^p)$ を支払う.
  - 集められた税は, 全市民に平等に分配される.
  - よって再分配後の市民 i の収入は,

$$\tilde{z}_i^p = z_i^p - T(z_i^p) + \frac{1}{N}\sum_{j=1}^{N} T\left(z_j^p\right) \tag{5}$$

  となる.

税ブラケット.

- 税額関数は, 次のように「ブラケット化」されたもののみを考える.
- カットオフ $\{m_b\}_{b=0}^{B}$ s.t. $0 = m_0 \leq m_1 \leq \cdots \leq m_{B-1} \leq m_B = +\infty$ が先に与えられる.
- 政府は, 各ブラケット b に含まれる収入に対して適用される限界税率 $\tau_b \in [0, 1]$ を選ぶことで, 税額 $T(\cdot)$ を決定する.

$$T(z) = \sum_{b=0}^{B-1} \tau_b \cdot ((m_{b+1} - m_b) \cdot 1[z > m_{b+1}] + (z - m_b) \cdot 1[m_b < z \leq m_{b+1}]) .$$

- 例:
  - カットオフ: $m_0 = 0,\ m_1 = 100, m_2 = 200, m_3 = 300, m_4 = +\infty.$
  - 限界税率: $\tau_0 = 0.1,\ \tau_1 = 0.2,\ \tau_2 = 0.3,\ \tau_3 = 0.4.$
  - 収入: $z_i^p = 250$
  - 最初の 100 には税率 0.1, 次の 100 には税率 0.2, 次の 50 には税率 0.3 がかかるので,

$$T(z_i^p) = 100 \times 0.1 + 100 \times 0.2 + 50 \times 0.3 = 45.$$

社会厚生関数

- 政府の目的関数である社会厚生関数は, 生産性と平等性のトレードオフを組み込めるように次のように決める.
- エージェントのコイン保有 $\boldsymbol{x}^c = (x_1^c, \ldots, x_N^c)$ に対し, 平等性を次で定義:

$$\mathbf{eq}(\boldsymbol{x}^c) = 1 - \mathbf{gini}(\boldsymbol{x}^c) \cdot \frac{N}{N-1}, \quad 0 \leq \mathbf{eq}(\boldsymbol{x}^c) \leq 1. \tag{7}$$

where

$$\mathbf{gini}(\boldsymbol{x}^c) = \frac{\sum_{i=1}^N \sum_{j=1}^N |x_i^c - x_j^c|}{2N \sum_{i=1}^N x_i^c}, \quad 0 \leq \mathbf{gini}(\boldsymbol{x}^c) \leq \frac{N-1}{N} \tag{8}$$

  - $\mathbf{eq}$ は, 1 で完全に平等 (全員同じ収入), 0 で完全に不平等 (1 人が全コインを独占).

- 生産性は,

$$\mathbf{prod}(\boldsymbol{x}^c) = \sum_{i=1}^{N} x_i^c. \tag{9}$$

- この平等性 **eq** と生産性 **prod** の積を社会厚生関数とする.

$$\mathbf{swf}_t(\boldsymbol{x}_t^c) = \mathbf{eq}_t(\boldsymbol{x}_t^c) \cdot \mathbf{prod}_t(\boldsymbol{x}_t^c). \tag{10}$$

政府の問題.

- 政府は,
  - 市民の保有資源・コイン $\boldsymbol{x}_{i,t}$, フィールドの状態（市民・資源の位置）と取引市場の状況は観測可能,
  - 各市民のスキル・蓄積労働コストは直接には観測できない.
- 政府の最大化問題は,

$$\max_{\pi_p} \mathbb{E}_{\tau\sim\pi_p, \boldsymbol{a}\sim\boldsymbol{\pi}, s'\sim\mathscr{T}} \left[\mathbf{swf}_0 + \sum_{t=1}^{H} \gamma^t \underbrace{(\mathbf{swf}_t - \mathbf{swf}_{t-1})}_{=r_{p,t}} \right]. \tag{12}$$

3.3 Inner-Outer-Loop Reinforcement Learning

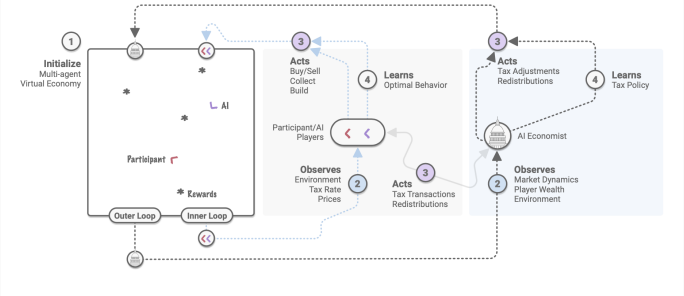


Figure 6: Two-level RL. In the inner loop, RL agents gain experience by performing labor, receiving income, and paying taxes, and learn through balancing exploration and exploitation how to adapt their behavior to maximize their utility. In the outer loop, the social planner adapts tax policies to optimize its social objective.

- この状況では, 内側のループで市民が行動を選んで学習し, 外側のループで政府が税制を選んで学習することになる.
- 市民の学習行動と政府の選ぶ税制が相互に影響し合うため, 報酬が不安定になり, 初期の学習に工夫が必要.

4.1 ベースライン

- 次の 4 つの税制度を比較する.
  - 無課税（フリーマーケット）
  - 2018 米連邦所得税
  - Saez フレームワーク
  - AI Economist
- ブラケットは全ての税制に共通で, 2018 年米所得税をもとに 1/1000 にスケーリングしたものを用いる.

$$\boldsymbol{m} = [0, 9.7, 39.475, 84.2, 160.725, 204.100, 510.3, \infty]. \tag{15}$$

2018 米連邦所得税.

- 2018 年の米連邦所得税をもとに, 各ブラケットの税率は,

$$\boldsymbol{\tau} = [0.1, 0.12, 0.22, 0.24, 0.32, 0.35, 0.37] \tag{16}$$

  とする.

Saez Tax Formula (single-period)

- Saez [2001] をもとに, まず single-period economy での optimal tax rates を求める.
- f, F は (pre-tax) 収入の分布の probability density と cummulative distribution function とし, planner はそれらを観測できるとする.
- Saez [2001] では, まず linear-weighted social welfare functions (11) に対し, social marginal welfare weights

$$g_i = \frac{\mathrm{d}\mathbf{swf}}{\mathrm{d}u_i}\frac{\mathrm{d}u_i}{\mathrm{d}x_i^c} = \omega_i \frac{\mathrm{d}u_i}{\mathrm{d}x_i^c}.$$

  を求める.
- これを我々のモデルに当てはめるために $g_i = \frac{1}{z_i}$ とし, さらにこれを normalize してよい $\sum_{i\in\mathscr{T}} g_i = 1$.

- $\alpha(z)$ を, the marginal average income at income z, normalized by the fraction of incomes above z, つまり,

$$\alpha(z) := \frac{z \cdot f(z)}{1 - F(z)}. \tag{18}$$

- $G(z)$ は, normalized, reverse cumulative Pareto weight over incomes above a threshold z

$$G(z) := \frac{1}{1 - F(z)} \int_{z'=z}^{\infty} f(z')g(z')\mathrm{d}z'. \tag{19}$$

- また, 弾力性 (elasticity) $e(z)$ を, average sensitivity of an agent's income to changes in the tax rate, つまり

$$e(z) = \frac{\mathrm{d}z/z}{\mathrm{d}(1 - \tau(z))/(1 - \tau(z))}. \tag{20}$$

- Saez [2001] によると, optimal marginal tax-rate は

$$\tau(z) = \frac{1 - G(z)}{1 - G(z) + \alpha(z)e(z)} \tag{21}$$

となり, これは income distribution と elasticity に依存して決まる.

Saez Tax Formula (multi-period)

- Saez formula (21) を使うには, income elasticity $e(z)$ の推定が必要.
- Gruber and Saez [2002] に従い, constant tax elasticity $\tilde{e}$ を仮定すると,

$$z_t = z^0 \cdot (1 - \tau_t)^{\tilde{e}}. \tag{22}$$

- したがって,

$$\log(z_t) = \tilde{e} \cdot \log(1 - \tau_t) + \log(z^0) \tag{23}$$

  を過去の tax period のデータから OLS で推定して用いる.

- AI Economist は, ディープニューラルネットワークにより各ブラケットの限界税率を決める.

$$\tau \sim \pi_p \left(o_{p,t}^{\text{world}}, o_{p,t}^{\text{agent}}, o_{p,t}^{\text{market}}, o_{p,t}^{\text{tax}}, h_{p,t-1}; \phi \right). \tag{24}$$

  - 基本のネットワークアーキテクチャは市民と同様 (Figure 3).
  - 市民と政府は観測できるものが異なる（政府はフィールド全体が見れるが, 市民のスキルは見れない）.

4.2 Training Strategy: Two-phase Training and Tax Curricula

- Section 3.3 で述べたように, この joint optimization は不安定性をもたらす.
- 学習を安定化させるため, ここでは次の two-pahse training approach を行った.
  - First phaze: agent models の集合に対し, 無課税 (no-tax scenario) で学習させる.
  - Second phaze: 税モデルを融合にし, エージェント・プランナーの学習を継続する. また突然の税導入による不安定性を回避するため, 限界税率の上限を 10% から 100% に徐々に引き上げる.
- また, planner policy に entropy regularization を行うことも効果的だった.
  - Entropy regularization は, policy gradient objective に次の additional, weighted term

    $$\mathbf{entropy}(\pi) = -\mathbb{E}_{a\sim\pi(\cdot|s)}[\log \pi(a \mid s)].$$

    を加える（Williams and Peng [1991], Mnih et al. [2016]）.

- 実験は RLlib flamework [Liang et al., 2018] を用いて実施.
- また Policy gradients の計算に, proximal policy gradients [Schulman et al., 2017] と Adam optimizer [Kingma and Ba, 2014] を使う.
- サンプルは 60 の environment replicas から平行して集め, sampling horizon は 200 timesteps とする.
- 全ての実験で, 400 million sample の phase two training を実施し, これは agent と planner model は stable policy への収束に十分であった.

4.3 平等性, 生産性と社会厚生

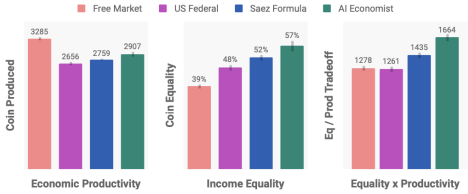


Figure 8: Comparison of overall economic outcomes (error bars show the variance during the final 20 million training steps, across 10 random seeds). The AI Economist achieves significantly better equality-productivity trade-offs compared to the baseline models. All baseline models have converged.

- 1 エピソード終了時の各税制での各指標の比較.
- 課税は常に生産性を下げるが, その下り幅は AI Economist が最も少ない.
- 平等性は, AI economist で最も高い.
- 社会厚生（平等性 $\times$ 生産性）も AI economist が最も高く, 次に高い Saez formula と比べても 16% 高い.

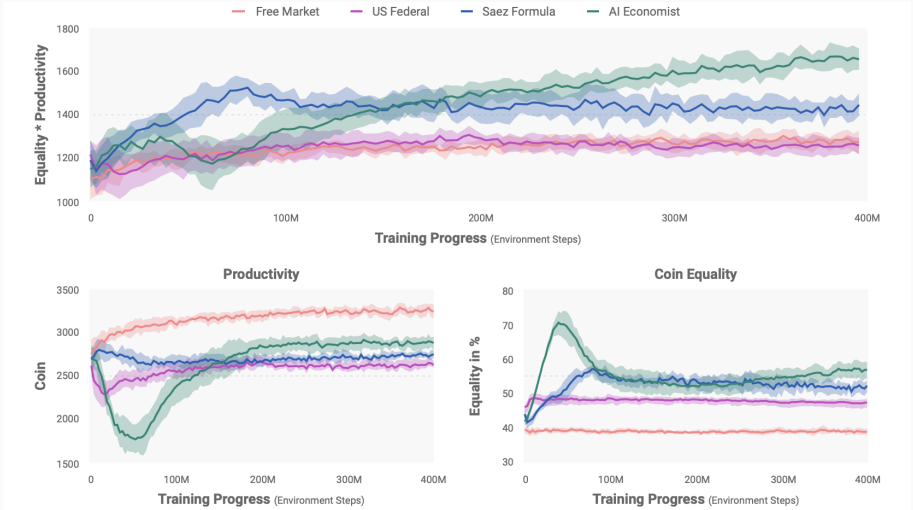


Figure 7: Empirical training progress for all models. The AI Economist (Green) achieves significantly better social outcomes than the baseline models. All baseline models have converged.

4.4 税額関数と課税額・再分配額の比較

税額関数

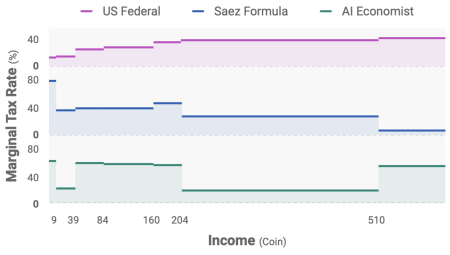


Figure 9: Comparison of average tax rates per episode. Variances within the Saez and AI Economist schedules are not shown. On average, the AI Economist sets a higher top tax rate than both of the US Federal and Saez tax schedules. The free-market collects zero taxes.

- 左図は各税制での限界税率の比較.
- 米所得税は高ブラケットに対して限界税率が上がっていく.
- Saez formula では概ね逆.
- AI Economist はそのブレンドになってる.

収入・課税額・再分配.

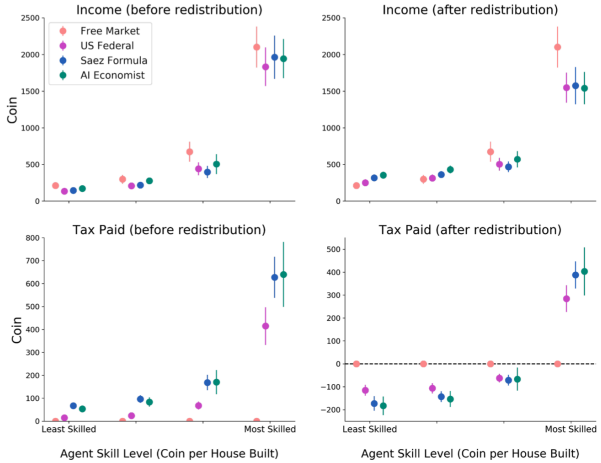


Figure 11: Agent-by-agent averages after sorting by skill. Income before redistribution (top-left) shows the average pre-tax income earned by each kind of agent. The amount of tax paid before distribution is shown in the bottom left. The amount of tax paid after redistribution is shown in the bottom right (the lower skill agents receive a net subsidy). The income after redistribution (top-right) shows the net average coin per agent at the

市民の行動への影響

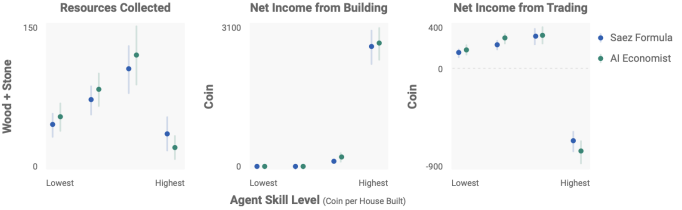


Figure 12: Comparing the impact of tax on economic activity for the Saez formula and the AI Economist. Left: Average number of resources (both wood and stone) collected per episode for each of the four agents. Middle: Average (per episode) total income earned from building. Right: Average (per episode) total income earned from trading. Negative income means the agent spends more than it earned. Lines indicate the standard deviation.

- 上手は, Saez framework と AI economist での市民の行動の比較.
- AI economist では, 高スキル市民は資源を集めるより, 市場で買って建築することに特化し, 低スキル市民はより資源を集めるようになっている.

4.5 Tax-Gaming Strategies

Figure 13: Left: Income of the highest-skilled agent for each tax period in an example episode with the AI Economist (green line). The dashed grey line shows the agent's average income. Right: Comparison of the total amount of tax the agent owed based on its actual income (green) and the tax it would have owed if it reported its average income in each period (gray). Each box in the column denotes the tax obligation in a single period.

- 上図は AI economist での高スキル市民の 1 エピソード間の収入推移.
- 高収入の税ピリオドと低収入の税ピリオドが交互に起きている.
- 市民は収入をばらつかせることで課税額を抑えるように行動している (Saez でも同様)

5. 被験者実験

- AI economist の税制が, 人間が参加する経済シミュレーションでも結果をよくするか検証する.
- Amazon Mechanical Turk (MTurk) platform において実験を行った.

5.1 実験方法

- 設定は基本的に同じだが, 若干の変更がある:
  - 取引は行わないようにする.
  - 労働コストがかかるのは建築のみとし, 移動・収集・取引は 0 コストとする. ただし建築の労働コストは 50% 増し.
  - 各エピソードは 5 分間で, フレーム率は 1 秒あたり 10 フレーム.
  - 各エピソードは 3000 ステップ. 各税ピリオドは 300 ステップ.

税制.

- 税制も AI 市民実験と基本的に同様.
- 無課税, 2018 米所得税はそのまま.
- Saez formula は, 学習後の 1 エピソード間の平均税率とする.
- AI economist は, AI 実験から効果的だった 1 制度を 1 つ使った (”Camelback” モデル)

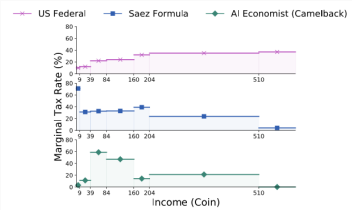


Figure 15: The "Camelback" model used in experiments with human participants. It features higher tax rates for incomes between 39 and 160 Coins compared to baselines.

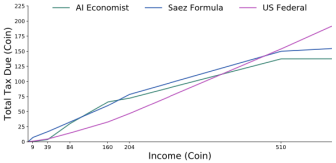


Figure 16: The effective taxes payable as a function of income under the "Camelback" schedule. The taxes grow faster under the Saez and AI Economist schedules. Note that these do not include the effect of subsidies. In effect, lower income workers receive net subsidies.

支払い.

- 各参加者はベース報酬 $5 とボーナス報酬最大 $10 を得る.
- このボーナスは,

$$\text{USD bonus} = \text{Utility} \times 0.06, \tag{26}$$

  で計算される.
- 平均の合計支払い額は, $11.26 だった.

5.2 結果

社会厚生の比較

- 社会厚生（平等性 × 生産性）は, AI economist の結果は米所得税, Saez フレームワークと比べて悪くなく, 無課税より有意に高かった.
- おおむね AI egents での実験と同等の結果.

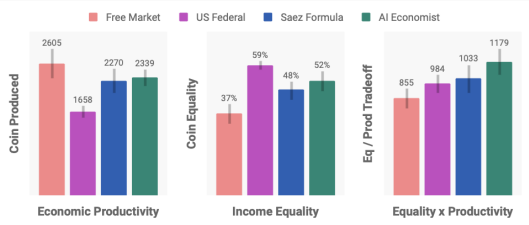


Figure 17: Social outcomes with 58 human participants in 51 episodes (first batch episodes with productivity of at least 1000 Coin). Each episode involves four participants. The AI Economist achieves competitive equality-productivity trade-offs with Saez and US Federal, and statistically significantly outperforms the free market (at $p = 0.05$). These results suggest a similar trend of improvement in equality-productivity trade-off as in the AI experiments.

6. 結論

- 生産性と平等性のトレードオフを上手くバランスする最適な税制を求めることは重要だが難しい.
- 経済シミュレーションにおいて, 深層強化学習を用いて政府主体に最適な税制を学習させることが考えられた.
- AI 市民と被験者による実験の双方で, Ai economist はベースライン税制（無課税・米所得税・Saez フレームワーク）と比較して高い水準の社会厚生を達成した.