Economics Bulletin

Volume 45, Issue 2

Misaligned beliefs and wealth distribution: An agent-based analysis

Wonjae Lee
Institute for Welfare State Research/Yonsei University

Bongkyoo Yoon
Korea National Defense University

Abstract

This study explores how individual mental models and beliefs—when misaligned with the objective structure of wealth distribution—can influence macroeconomic inequality. Using an integrated framework that combines agent-based modeling (ABM) and system dynamics (SD), we investigate the effects of micro-level behavioral assumptions on wealth outcomes. Drawing inspiration from Michael Sandel's critique of meritocracy, we test whether cognitive biases, such as those grounded in prospect theory and meritocratic perceptions, can reshape wealth dynamics even in a purely stochastic economic environment. Our simulations are based on a modified version of Wilensky's Simple Economy model, termed the Flexible Simple Economy, which introduces agents with differing risk profiles based on behavioral economics. The findings reveal that mental models significantly affect the long-term distribution of wealth: prospect theory-based agents produce more egalitarian outcomes, whereas meritocratic beliefs tend to exacerbate inequality. These results suggest that policies targeting belief systems and behavioral expectations—rather than solely redistributive interventions—may be crucial for addressing economic inequality. By demonstrating a feedback loop between individual perceptions and systemic outcomes, this study highlights the central role of cognitive and psychological factors in shaping economic distributions.

Citation: Wonjae Lee and Bongkyoo Yoon, (2025) "Misaligned beliefs and wealth distribution: An agent-based analysis", *Economics Bulletin*, Volume 45, Issue 2, pages 763-779
Contact: Wonjae Lee - timelast@naver.com, Bongkyoo Yoon - bkyoon1@gmail.com.
Submitted: June 18, 2024. **Published:** June 30, 2025.

Misaligned Beliefs and Wealth Distribution: An Agent-Based Analysis

1. Introduction

Michael Sandel's (2021) critique of meritocracy in his book "The Tyranny of Merit" has sparked a significant social debate. Sandel argues that the success of the American upper and middle classes is often falsely attributed to individual merit when it is largely due to luck or heritage. He contends that an excessive belief in meritocracy reinforces overinvestment in education, leading to greater inequality. While this argument is worthy of discussion in its own right, it raises an important and under-explored scientific question: Can micro-level "beliefs" held by individuals systematically change the distribution of wealth in the market, leading to macro-level outcomes of inequality, even when these beliefs are misaligned with reality?

Wealth distribution studies have primarily focused on uncovering general laws governing the distribution of wealth from empirical data. Pareto (1897) pioneered the search for these laws, finding that income followed a specific, heavily skewed distribution toward the top in most cities at the end of the 19th century. Subsequent studies have further explored the nature of wealth distribution, with some finding exponential (Drăgulescu & Yakovenko, 2001; Tao et al., 2019) or Pareto distributions (Jones, 2015) based on empirical data. However, these studies have not investigated the influence of individual mental models on the market mechanism itself and the resulting inequality.

Behavioral economics research, on the other hand, has extensively used experiments to explain how people develop mental models of reality and how these models drive behavior that may be misaligned with reality. Prospect theory (Kahneman & Tversky, 1979) and various biased behavioral economic properties, such as the endowment effect, loss aversion, and status quo bias (Kahneman et al., 1991), have been introduced to explain seemingly irrational decision-making. Kahneman (2003) synthesizes these concepts into the notion of "bounded rationality." However, these studies do not typically examine whether these behavioral economic properties lead to specific macro-level changes, such as alterations in the distribution of wealth.

Some research has sought to explore the interaction between individual beliefs and market mechanisms. Solt et al. (2016) find that higher levels of economic inequality can strengthen beliefs in meritocracy, even when such beliefs may not be justified by actual economic mobility. Cruces et al. (2013) demonstrate that biased perceptions of income distribution affect preferences for redistribution, highlighting the feedback loop between beliefs and economic policies. Cappelen, de Haan, and Tungodden (2024) delve into how individuals process limited information to form beliefs about fairness and meritocracy, arguing that people tend to be "Bayesian meritocrats." While these studies provide valuable insights into the relationship between inequality structures and individual perceptions, they do not directly address how mental models themselves can affect the distribution of wealth.

To fill this research gap, we conduct a hypothetical thought experiment by narrowing down our research question: In a society where wealth distribution is solely dependent on luck, would the distribution of wealth differ when participants with specific beliefs about the distribution rules compete, compared to a situation without such beliefs? For example, in this hypothetical society, whether the belief is meritocratic or its opposite, it may not increase the overall wealth of the society or enable individuals to acquire more or less wealth. However, could it still cause changes in the distribution of wealth? The purpose of this study is to test these intuitive hypotheses using agent-based modeling (ABM) and system dynamics (SD) models.

In a broader sense, this research aims to explore whether transformations in mental models within civil society can trigger systemic changes in economic distribution. If they can, it would necessitate fundamental changes in policy measures to address economic inequality. This study intends to contribute to this crucial discussion.

2. Methodology

To bridge the research gap between behavioral economics and wealth distribution studies, we propose an integrated modeling approach that combines system dynamics (SD) and agent-based modeling (ABM). This approach allows us to incorporate individual mental models, as studied in behavioral economics, into the analysis of wealth distribution dynamics at the macro level.

The methodological divergence between behavioral economics and wealth distribution studies is the main reason for the identified research gap. In behavioral economics, research on mental models is primarily conducted through experiments on individual behavior in controlled environments. These studies provide a detailed understanding of how individual mental models, or psychological processes, distort reality, create biases, and ultimately lead to irrational decisions. However, in these studies, the macro-level wealth distribution is often treated as an exogenous variable. In contrast, most wealth distribution studies rely on large-scale administrative or survey data to examine the distribution of income or assets among individuals, treating individuals as atoms without mental models. Consequently, while the general laws about wealth distribution and individual mental models have been studied in detail, the interaction between the two is difficult to describe. The research gap arises precisely from the methodological heterogeneity between macro-level wealth distribution studies and micro-level mental model studies.

To address this gap, we employ system dynamics (SD) to model individual mental models, capturing the dynamic decision-making process at the micro level. Although SD is commonly known for its usefulness in modeling macro-level system dynamics, it is also a valuable tool for modeling individual mental processes (Sterman, 2001). SD allows us to represent mental models where the inflow is real-world information that an individual encounters, the stock is the accumulation of the information, and the outflow is the behavior or decisions originating from the inflow and the stock. Concurrently, at the macro level, we use agent-based modeling (ABM) to simulate a market where agents interact and engage in transactions. ABM is suitable for analyzing how micro-behaviors (actions of individuals or small groups) influence macro-outcomes (the state of the entire system) (Wilensky & Rand, 2015). This design allows us to examine the feedback loop between individual decision-making processes and market functioning.

Our integrated modeling approach consists of the following key steps:
(1) Designing the Market: Implement the market structure and rules in an agent-based modeling (ABM) framework, where agents interact in a simulated economy.
(2) Designing Agents: Use a system dynamics framework to develop individual agents' mental models that are misaligned with reality but guide their decision-making within the simulated market, based on insights from behavioral economics literature.
(3) Designing the Integrated Model: Link the agents to the market, using their market

performance as an inflow to their mental models and their risk-taking decisions as an outflow back to the market.
(4) Running Experiments: Simulate various economic scenarios with different agents' mental models to analyze the impact of individual mental models on wealth distribution dynamics.

3. Model

3.1. Market: The Flexible Simple Economy

The Flexible Simple Economy model, a modified version of Wilensky and Rand (2015)'s Simple Economy model, serves as the market model for this study. The Simple Economy model simulates a discrete-time closed economy where agents engage in wealth transactions purely based on chance. Despite its simplicity, this model remarkably replicates the exponentially distributed wealth patterns observed in real-world economies (Wilensky, 1998; Tao et al., 2019).

When N denotes the total population and $W_t^{(i)}$ denotes the wealth which agent i owns at time t, Simple Economy Model Rules are:
(1) The population consists of $(N = 500)$ agents.
(2) Each agent initially possesses $(W_0^{(i)} = \$100)$ in wealth.
(3) At each time step t, an agent i randomly gives \$1 to another agent j, where $j \neq i$.
(4) No agent's wealth $W_t^{(i)}$ can fall below zero, i.e., $(\forall i, W_t^{(i)} \geq 0)$.
(5) Agents who run out of money $W_t^{(i)}$=0 cannot give money to others until they have money again.
(6) The total amount of money in the system is conserved $(\forall_{i,t}, \sum_i W_t^{(i)} = N \times W_0^{(i)})$.

As a result, the wealth that agent i owns at time t is the wealth after N-th transaction – the last transaction for the time $(W_t^{(i)} = W_{t,N}^{(i)})$. Therefore, the wealth an agent i holds at time t before the initiation of any transaction is equivalent to the wealth after the completion of all transactions at time $t-1$ $(W_{t,0}^{(i)} = W_{t-1,N}^{(i)})$.

Now, we define $D_{t,k}^{(i\to j)}$ as the amount of wealth transferred from agent i to j during the k-th transaction at time t, as noted in the rule (3) above. This transfer function is given by:

$$D_{t,k}^{(i\to j)} = \begin{cases} 1, & i \neq j \text{ } and \text{ } W_{t,k-1}^{(i)} \geq 1 \\ 0, & otherwise \end{cases}. \tag{1}$$

Since $D_{t,k}^{(j\to i)}$ is the amount of wealth transferred from agent j to i after the k-th transaction at time t, the wealth of agent i is updated as follows:

$$\begin{aligned} W_{t,k}^{(i)} &= W_{t,k-1}^{(i)} - D_{t,k}^{(i\to j)} + \sum_{j\neq i} D_{t,k}^{(j\to i)} \\ &= W_{t,0}^{(i)} - \sum_{p=1}^{k} D_{t,p}^{(i\to j)} + \sum_{p=1}^{k}\sum_{j\neq i} D_{j,p}^{(j\to i)}. \end{aligned} \tag{2}$$

In the Flexible Simple Economy model, rule (3) is modified into rule (3-1) with Equation (3) below to allow transaction patterns to vary based on the agent's risk tolerance level:

(3-1) At each time step t, agents are categorized as risk-seeking or risk-averse based on their tolerable risk level. A risk-seeking agent i gives \$2 to another risk-seeking agent j (where $j \neq i$), while a risk-averse agent i gives \$1 to another risk-averse agent j (where $i \neq j$). Therefore, Equation (1) is modified as follows:

$$D_{t,k}^{(i \to j)} = \begin{cases} 2, & i \neq j,\ W_{t,k-1}^{(i)} \geq 2, and\ agent\ i\ and\ j\ are\ risk-seeking \\ 1, & i \neq j, W_{t,k-1}^{(i)} \geq 1, and\ agent\ i\ and\ j\ are\ risk-averse \\ 0, & otherwise. \end{cases} \tag{3}$$

This flexibility in the model allows for the observation of changes in the market system resulting from variations in agents' mental models. The modified rule introduces heterogeneity in the agents' risk preferences, which influences their trading behavior. Risk-seeking agents engage in higher-stakes transactions, transferring larger amounts of wealth per interaction, while risk-averse agents opt for smaller, more conservative transfers.

It is important to note that in this market model, an agent's risk preference affects their individual trade decisions but does not necessarily impact their average performance in terms of wealth accumulation. The model is designed such that the expected value of wealth transfer remains the same for both risk-seeking and risk-averse agents. This design choice allows for a focused examination of how variations in risk preferences and trading behavior can influence the overall distribution of wealth in the system, without confounding the results with differences in average agent performance.

3.2. Agents

3.2.1. Prospect Theory-Based Agent

Kahneman and Tversky's (1979) prospect theory posits that individuals evaluate economic outcomes relative to a reference point, which separates potential gains from losses and impacts sensitivity to gains and losses differently. Specifically, individuals exhibit loss aversion, being more sensitive to losses than to gains above the reference point, and risk-seeking behavior when below the reference point.

Note that both mental models are misaligned with reality, as the agents adjust their investment decisions based on past performance or current wealth, even though neither past performance nor wealth directly affects future investment outcomes in our model, where all transactions are performed on a purely random basis.

For simplicity, we define the prospect theory-based agent as follows:

(1) Risk Aversion Above the Reference Point:
Agent i at time t, whose current wealth is greater than the initial wealth ($W_t^{(i)} > W_0^{(i)}$) perceives their 'prospect' as negative, becoming risk-averse and preferring lower investment amounts.

(2) Risk Seeking Below the Reference Point:
Agent i at time t, whose current wealth is the same as or less than the initial wealth ($W_t^{(i)} \leq W_0^{(i)}$) perceives their 'prospect' as positive, becoming risk-seeking and preferring higher investment amounts.

3.2.2. Meritocratic Agent

We design a meritocratic agent using Chatterjee et al. (2011)'s research on capability cues. The study suggests that CEOs tend to assess their own abilities based on environmental changes, such as recent profit levels of their firms. If these cues suggest high capability, CEOs are inclined towards high-risk investment strategies; conversely, if the cues suggest low capability, they prefer low-risk investment strategies. A mental model with capability cues reflects the very nature of meritocracy, in that it is a belief that one's performance is solely dependent upon one's capabilities.

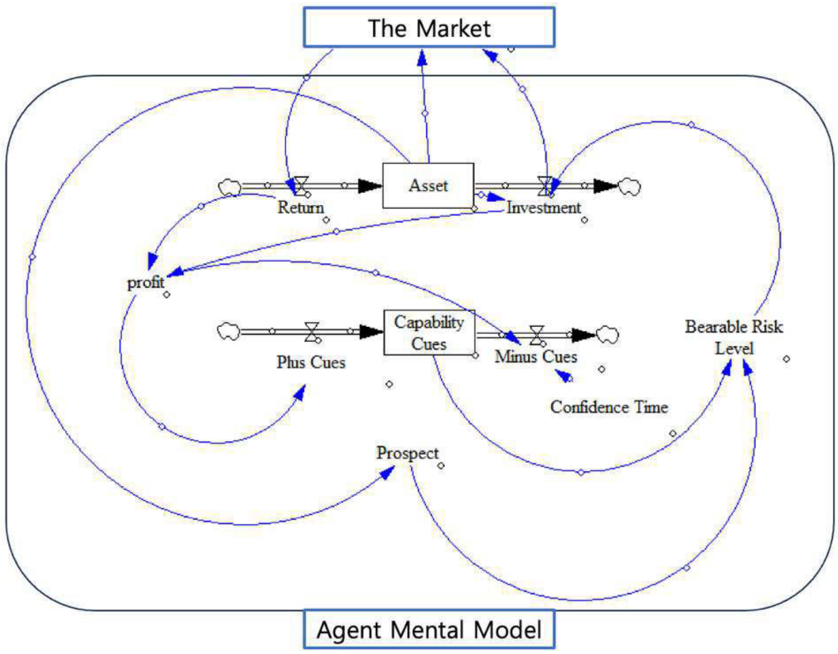
For simplicity, we define the meritocratic agent as follows:

(1) Risk Aversion with Negative Capability Cues:
- Agent i for time t, for whom the sum of return over the past five periods is less than or equal to the sum of their investments ($\sum_{t-4}^{t} R_t^{(i)} \leq \sum_{t-4}^{t} I_t^{(i)}$) perceive their 'capability cue' as negative, leading to risk-averse behavior and a preference for lower investment amounts.

(2) Risk Seeking with Positive Capability Cues:
- Agent i for time t, for whom the sum of return over the past five periods exceeds the sum of their investments ($\sum_{t-4}^{t} R_t^{(i)} > \sum_{t-4}^{t} I_t^{(i)}$) perceive their 'capability cue' as positive, leading to risk-seeking behavior and a preference for higher investment amounts.

Note that there are delays between performance itself and the resulting behavior change. Sterman (2001) correctly modeled system dynamics mental models, accounting for delays between reality and perception, illustrating how such gaps can transform the entire system. The feedback loops between the agent mental model and the market are depicted in Figure 1.

Figure 1. Feedback Loop of the Flexible Simple Economy



In Figure 1, “investment” is $D_{t,k}^{(i\to j)}$, “return” is $\sum_{j\neq i} D_{t,k}^{(j\to i)}$, and “bearable risk level” indicates whether agent i is risk-seeking or risk-averse at time t. Agents take returns from the market as inflows and feed back investments to the market as outflows according to their own mental models, forming a feedback loop for the Flexible Simple Economy.

3.3. Baseline and Comparative Models

3.3.1. Baseline Model: Simple Economy

The original Simple Economy model without agent mental models serves as the baseline model for this study. This baseline model is effectively equivalent to the Flexible Simple Economy where all agents are risk-averse. The baseline model is compared with two comparative models, in which the market operates entirely randomly, but the agents, unaware of this randomness, create rational mental models to make investment decisions. The study then observes how wealth distribution changes in comparison to the baseline model.

3.3.2 Comparative Model A: Moderate Rich's Society

Model A is the Flexible Simple Economy with prospect theory-based agents. These agents use their initial asset size as the reference point. They perform high-risk investing when their asset sizes are above the reference points and low-risk investing when their asset sizes are equal to or below the reference points.

3.3.3 Comparative Model B: Rise of Meritocracy

Model B is the Flexible Simple Economy with meritocratic agents. Although the model operates purely by chance, agents mistakenly believe they are in a meritocratic society and make decisions based on capability cues. Agents with positive capability cues make risk-seeking investments, while those with negative cues make risk-averse decisions.

By comparing these models, the study aims to understand the impact of different mental models on wealth distribution within a purely random market environment.

4. Results

4.1 Evaluation Metrics

This study adopts Lee and Yoon (2023)'s proposal of four evaluation metrics as the evaluation criteria for analyzing wealth distribution.

(1) Distribution shape: The distribution is considered more unequal if it fits an exponential distribution more quickly, as the steady-state distribution of the Simple Economy is exponential.
(2) Gini coefficient: A higher Gini coefficient indicates greater inequality.
(3) Share of wealth held by the top 1% of agents: A larger share of wealth held by the top 1% of agents suggests greater inequality.
(4) Share of wealth held by the top 10% of agents: Similarly, a larger share of wealth held by the top 10% of agents indicates greater inequality.

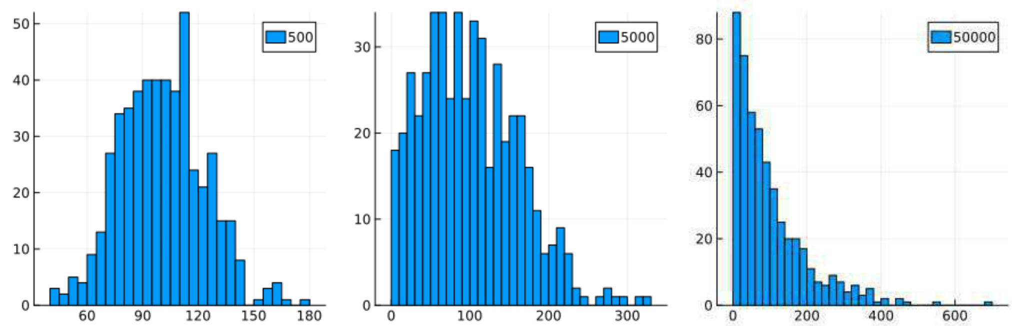
4.2 Changes in Wealth Distribution

The baseline model, Simple Economy, starts with a uniform distribution where each agent initially has $100. Over the long term (50,000 time steps), it evolves into an exponential distribution as Wilensky & Rand (2015) described, with many agents holding low wealth and a few holding high wealth (Figure 2). To capture perception delays, we assume a memory length of 5 turns for agents' perceived capability in the meritocratic scenario, meaning agents perceive their capability based on average performance over the last five turns.

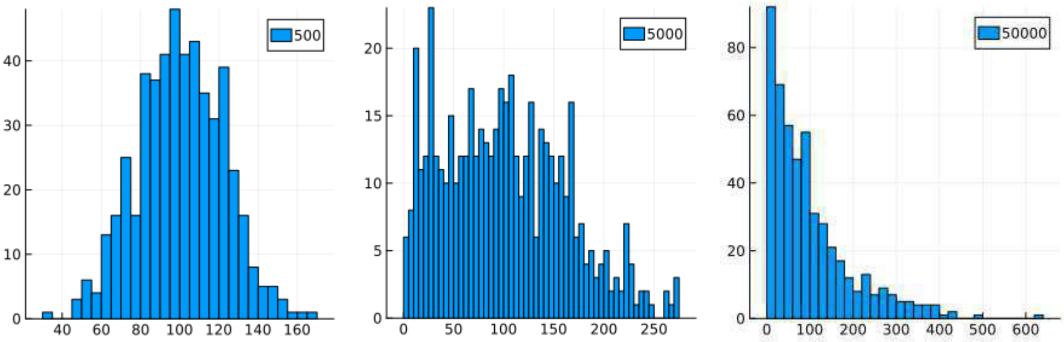
While the base case assumes a 5-turn memory and initial wealth of $100, sensitivity analyses are performed with memory lengths of 8 and 10 turns and initial wealth levels of $80 and $120. The resulting distributions over time under these alternative conditions are presented in Figures 2 to 4.

Figure 2: Baseline model wealth distributions at turn 500, 5000, and 50000.

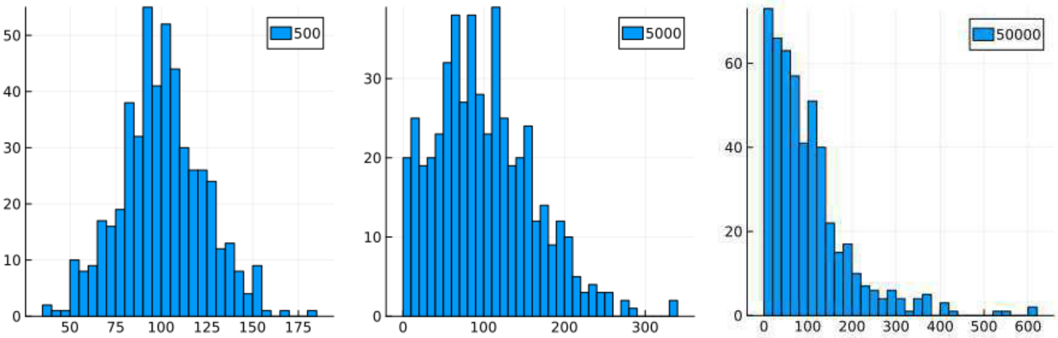
<memory = 5, initial wealth = 100>



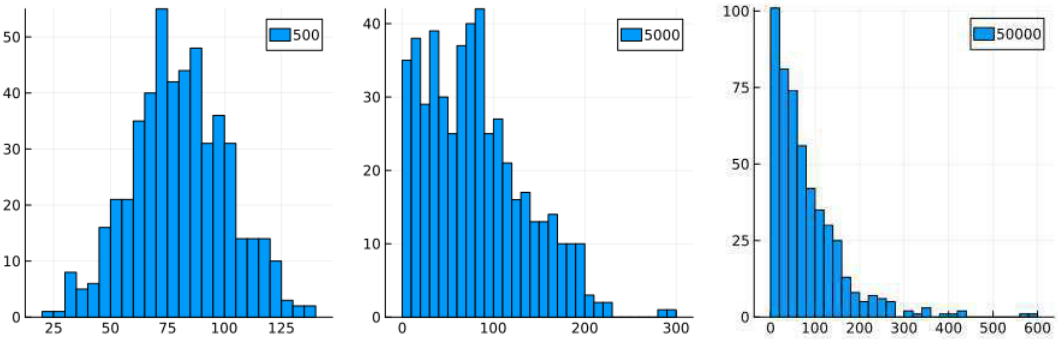
<memory = 8, initial wealth = 100>



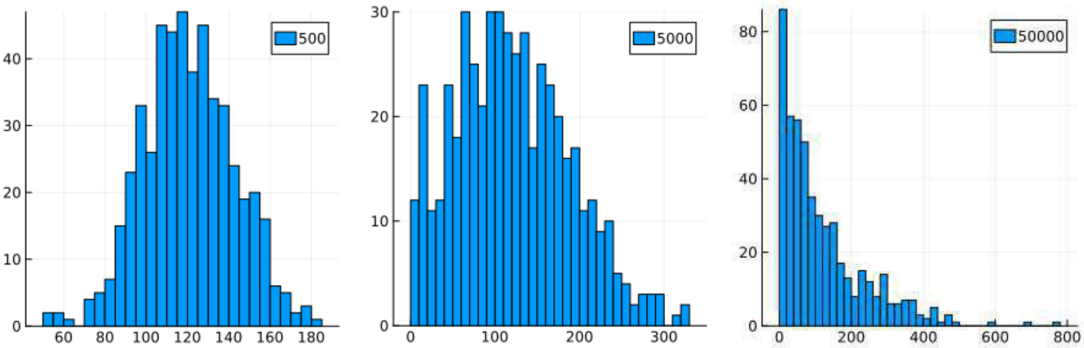
<memory = 10, initial wealth = 100>



<memory = 5, initial wealth = 80>



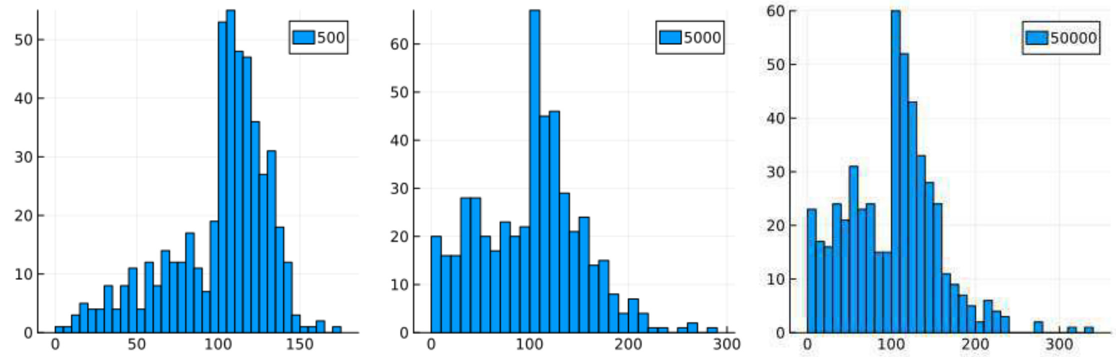
<memory = 5, initial wealth = 120>



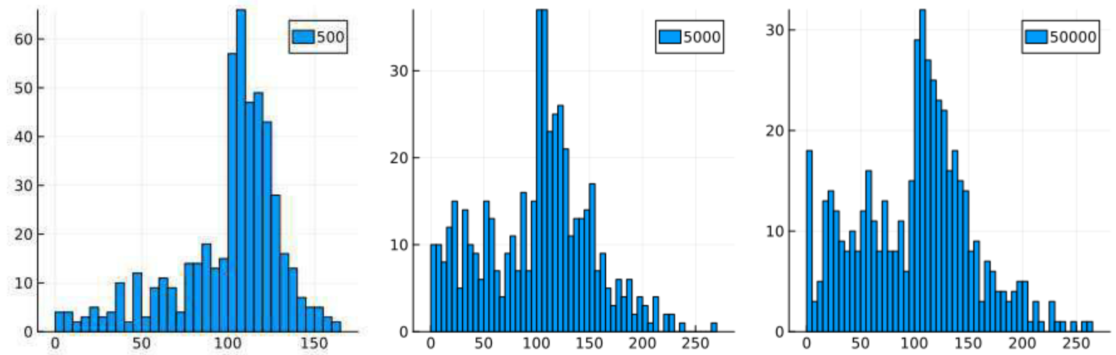
Comparative Model A, which also begins with a uniform distribution, maintains a more even distribution over time, with the middle tier being the most populous (Figure 3). Unlike the baseline model, Comparative Model A does not trend towards an exponential distribution.

Figure 3: Model A’s wealth distributions at turn 500, 5000, and 50000.

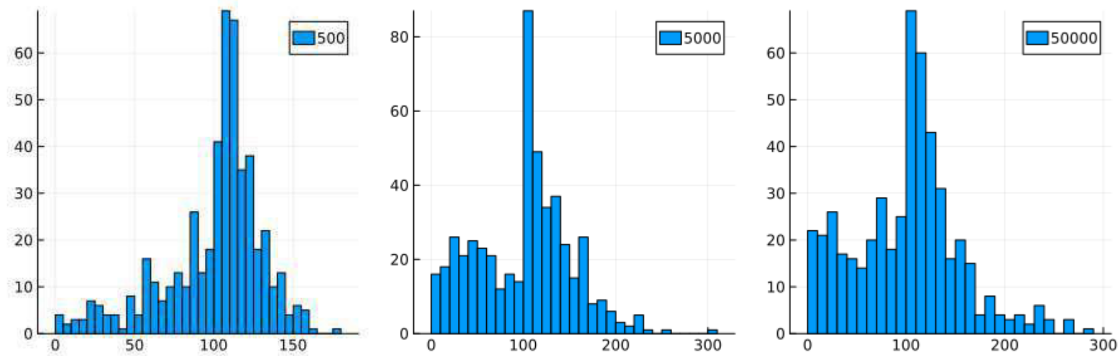
<memory = 5, initial wealth = 100>



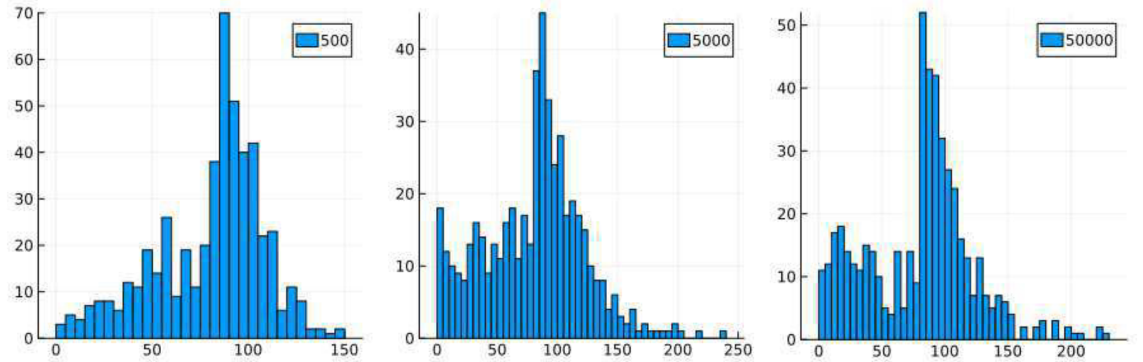
<memory = 8, initial wealth = 100>



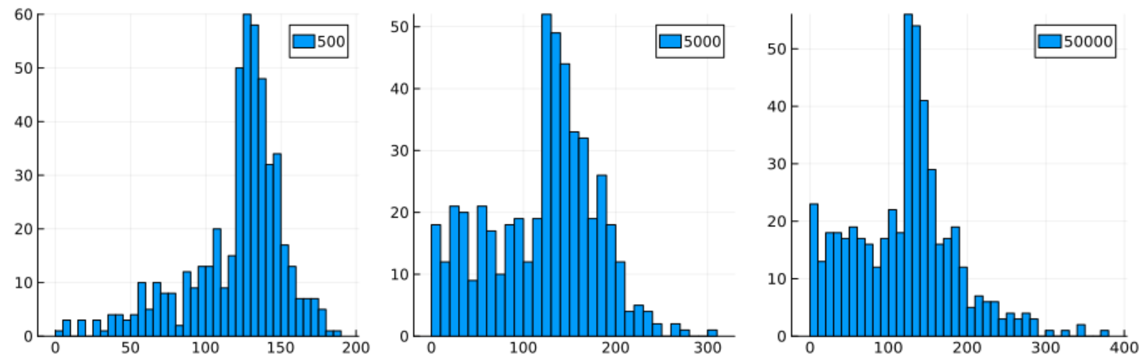
<memory = 10, initial wealth = 100>



<memory = 5, initial wealth = 80>



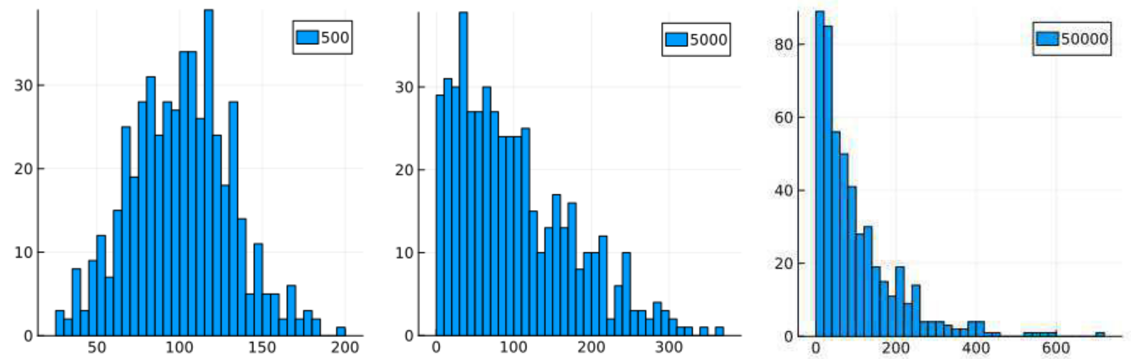
<memory = 5, initial wealth = 120>



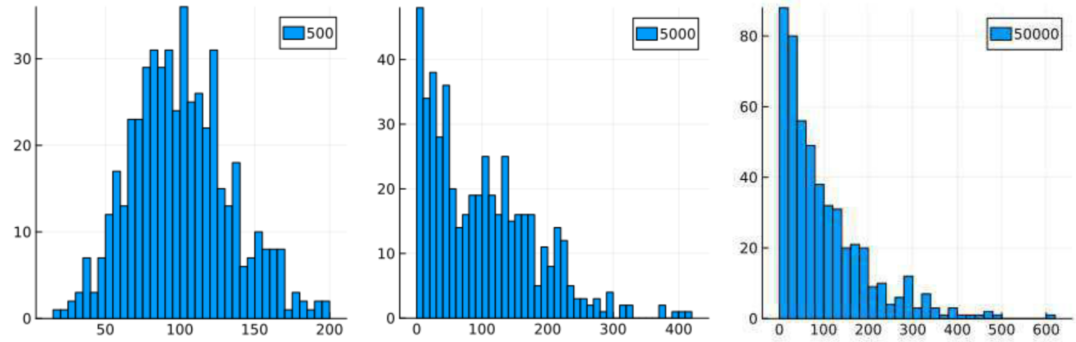
Comparative Model B, however, beginning with a uniform distribution, follows a similar long-term exponential distribution pattern as the baseline model (Figure 4).

Figure 4: Model B's wealth distributions at turn 500, 5000, and 50000.

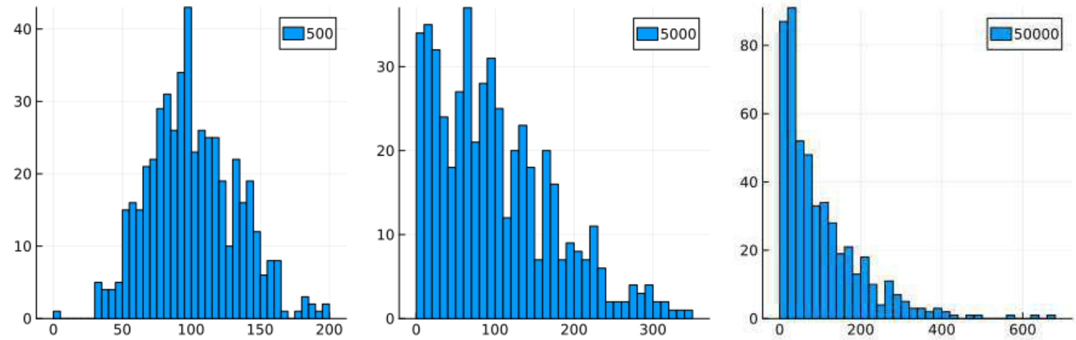
<memory = 5, initial wealth = 100>



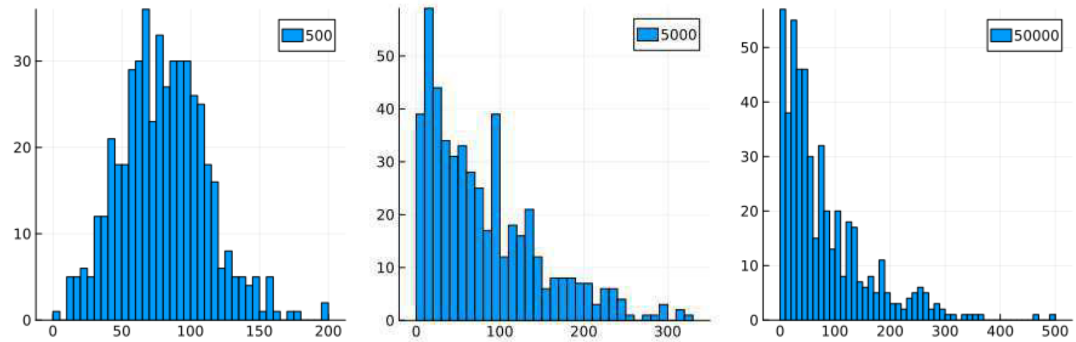
<memory = 8, initial wealth = 100>



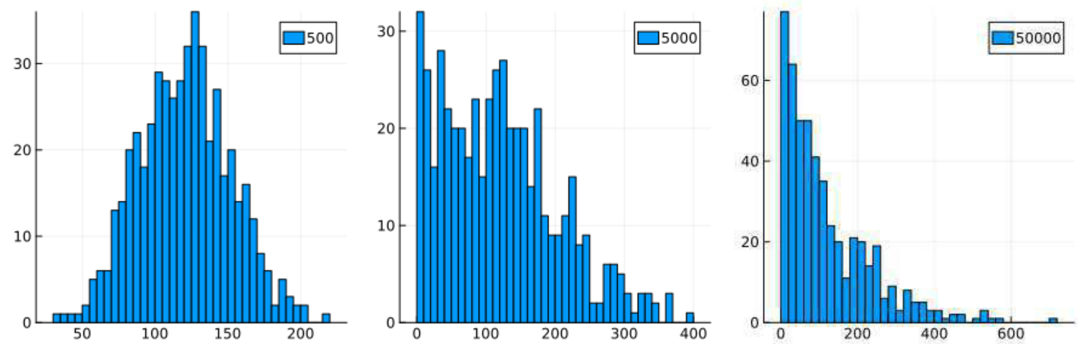
<memory = 10, initial wealth = 100>



<memory = 5, initial wealth = 80>



<memory = 5, initial wealth = 120>



To evaluate the wealth distribution models, we conducted a Kolmogorov-Smirnov test for each model at turn 50,000 (Table 1). To control for stochastic effects inherent in agent-based simulations, we utilize an ensemble average of 100 simulations for statistical robustness in the

Kolmogorov-Smirnov (K-S) tests. Silveira et al. (2018) demonstrated that ensemble-based approaches significantly enhance the statistical robustness of Kolmogorov-Smirnov tests compared to single-simulation applications, with approximately 100 ensemble members representing the optimal balance between stochastic control and computational efficiency.

The decision to accept or reject the null hypothesis (H_0) for the K-S tests is based on two steps: firstly, the median p-value across simulations must exceed the 0.05 threshold; secondly, at least 80% of the simulations must individually fail to reject H_0, thereby ensuring robustness and consistency in hypothesis testing results (Lorscheid, Heine, & Meyer, 2012).

Table 1. Ensemble Averages of Kolmogorov-Smirnov Test Results for Wealth Distribution Models (number of simulations = 100, initial wealth = 100)

Model	Acceptance case (p-value>0.05)	H_0	p-value (median)	Outcome
Baseline	89%	Exponential Distribution	0.2624	Fail to reject H_0
Baseline	32%	Normal Distribution	0.0142	Reject H_0
Model A	1%	Exponential Distribution	<1e-05	Reject H_0
Model A	81%	Normal Distribution	0.2577	Fail to reject H_0
Model B	94%	Exponential Distribution	0.3096	Fail to reject H_0
Model B	33%	Normal Distribution	0.0163	Reject H_0

The Baseline Model did not reject the null hypothesis for an exponential distribution (median p-value = 0.2624), indicating it closely follows an exponential pattern. Model A rejected the null hypothesis for an exponential distribution (median p-value < 1e-05), suggesting significant deviation, but did not reject it for a normal distribution (median p-value = 0.2577). Model B, like the Baseline Model, did not reject the null hypothesis for an exponential distribution (p-value = 0.3096), but rejected it for a normal distribution (p-value = 0.0163), indicating a deviation towards an exponential distribution. These results suggest that Model A (Moderate Rich's Society) produced a long-term wealth distribution distinct from that of the baseline Simple Economy model, whereas Model B (Rise of Meritocracy) did not. This highlights the significance of agent mental models in shaping long-term wealth distribution.

Tables 2 and 3 present the detailed K-S test results from sensitivity analyses conducted on each model, varying perception delays and initial wealth levels. Note that the baseline case (memory = 5, initial wealth = 100) represents the ensemble average of 100 simulations.

Table 2. Perception Delays Sensitivity in KS Test Results (sample size = 100, Initial Wealth = 100)

Model	Memory (delay)	H_0	p-value	Outcome (% of p-value>0.05 case)
Baseline Model (100 simulations)	5	Exponential Distribution	0.2624 (median)	Fail to reject H_0 (89%)
Baseline Model	6	Exponential Distribution	0.1141	Fail to reject H_0
Baseline Model	8	Exponential Distribution	0.0292	Reject H_0
Baseline Model	10	Exponential Distribution	0.2890	Fail to reject H_0
Baseline Model (100 simulations)	5	Normal Distribution	0.0142 (median)	Reject H_0 (32%)
Baseline Model	6	Normal Distribution	0.0175	Reject H_0
Baseline Model	8	Normal Distribution	0.1666	Fail to reject H_0
Baseline Model	10	Normal Distribution	0.2213	Fail to reject H_0
Model A	5	Exponential Distribution	<1e-05	Reject H_0

(100 simulations)			(median)	(1%)
Model A	6	Exponential Distribution	<1e-07	Reject H_0
Model A	8	Exponential Distribution	<1e-05	Reject H_0
Model A	10	Exponential Distribution	<1e-07	Reject H_0
Model A (100 simulations)	5	Normal Distribution	0.2577 (median)	Fail to reject H_0 (81%)
Model A	6	Normal Distribution	0.4542	Fail to reject H_0
Model A	8	Normal Distribution	0.3939	Fail to reject H_0
Model A	10	Normal Distribution	0.4044	Fail to reject H_0
Model B (100 simulations)	5	Exponential Distribution	0.3096 (median)	Fail to reject H_0 (94%)
Model B	6	Exponential Distribution	0.6847	Fail to reject H_0
Model B	8	Exponential Distribution	0.0358	Reject H_0
Model B	10	Exponential Distribution	0.3706	Fail to reject H_0
Model B (100 simulations)	5	Normal Distribution	0.0163 (median)	Reject H_0 (67%)
Model B	6	Normal Distribution	<0.0314	Reject H_0
Model B	8	Normal Distribution	0.0523	Fail to reject H_0
Model B	10	Normal Distribution	<1e-04	Reject H_0

Table 3. Initial Wealth Sensitivity in KS Test Results (sample size = 100, memory = 5)

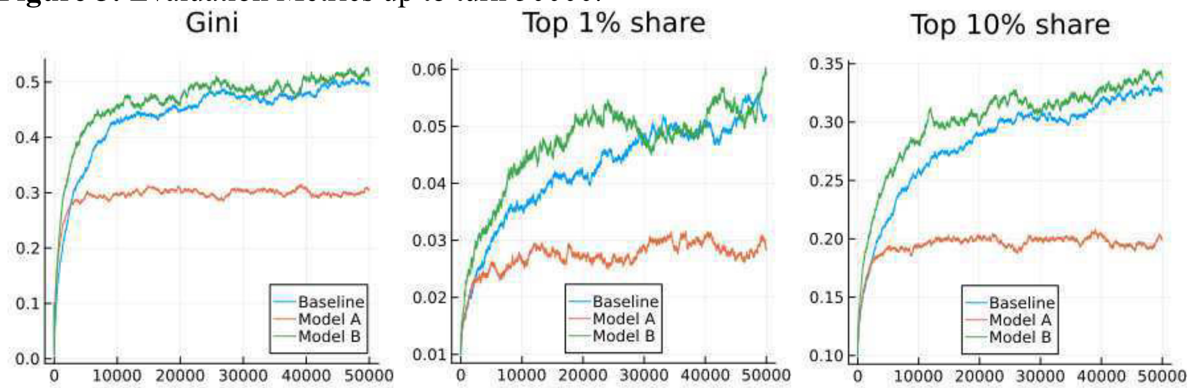
Model	Initial Wealth	H_0	p-value	Outcome (95% Confidence)
Baseline Model	80	Exponential Distribution	0.2290	Fail to reject H_0
Baseline Model	90	Exponential Distribution	0.1271	Fail to reject H_0
Baseline Model (100 simulations)	100	Exponential Distribution	0.2624 (median)	Fail to reject H_0 (89%)
Baseline Model	110	Exponential Distribution	0.6762	Fail to reject H_0
Baseline Model	120	Exponential Distribution	0.1033	Fail to reject H_0
Baseline Model	80	Normal Distribution	0.0800	Fail to reject H_0
Baseline Model	90	Normal Distribution	0.0002	Reject H_0
Baseline Model (100 simulations)	100	Normal Distribution	0.0142 (median)	Reject H_0 (32%)
Baseline Model	110	Normal Distribution	0.0481	Reject H_0
Baseline Model	120	Normal Distribution	0.0004	Reject H_0
Model A	80	Exponential Distribution	<1e-05	Reject H_0
Model A	90	Exponential Distribution	<1e-04	Reject H_0
Model A (100 simulations)	100	Exponential Distribution	<1e-05 (median)	Reject H_0 (1%)
Model A	110	Exponential Distribution	<1e-05	Reject H_0
Model A	120	Exponential Distribution	<1e-09	Reject H_0
Model A	80	Normal Distribution	0.0248	Reject H_0
Model A	90	Normal Distribution	0.4710	Fail to Reject H_0
Model A (100 simulations)	100	Normal Distribution	0.2577 (median)	Fail to reject H_0 (81%)
Model A	110	Normal Distribution	0.3381	Fail to reject H_0
Model A	120	Normal Distribution	0.6691	Fail to reject H_0
Model B	80	Exponential Distribution	0.5829	Fail to reject H_0
Model B	90	Exponential Distribution	0.0957	Fail to reject H_0
Model B (100 simulations)	5	Exponential Distribution	0.3096 (median)	Fail to reject H_0 (94%)

Model B	110	Exponential Distribution	0.0883	Fail to reject H_0
Model B	120	Exponential Distribution	0.4878	Fail to reject H_0
Model B	80	Normal Distribution	0.0001	Reject H_0
Model B	90	Normal Distribution	0.0348	Reject H_0
Model B	100	Normal Distribution	0.0163	Reject H_0
Model B	110	Normal Distribution	0.0107	Reject H_0
Model B	120	Normal Distribution	0.0038	Reject H_0

The sensitivity analysis shows that the ensemble average outcomes in the baseline scenario are largely supported across varying parameterizations. Specifically, despite slight variations—such as a rejection of the exponential hypothesis at an 8-turn memory and minor discrepancies in single simulation results when initial wealth shifts from $100—the core pattern of an exponential wealth distribution remains robust. These findings affirm that the ensemble average (based on 100 simulations) reliably captures the underlying dynamics of wealth distribution, even as individual simulation outcomes exhibit modest variability.

Other evaluation metrics support the distinct wealth distribution from the two comparative models. In Model A, involving prospect theory-based agents, the Gini coefficient (Figure 5, left), the share of wealth held by the top 1% (Figure 5, center), and the share of wealth held by the top 10% (Figure 5, right) all fall below the levels observed in the baseline model. In contrast, Model B, involving capability cues-based agents, exhibits higher levels than the baseline model for the Gini coefficient, top 1% wealth share, and top 10% wealth share. Higher values for the Gini coefficient and the wealth shares of the top 1% and top 10% indicate greater inequality.

Figure 5: Evaluation Metrics up to turn 50000.



5. Final Remarks

Our findings demonstrate that the mental models adopted by agents can significantly influence the distribution of wealth, even in a market where transactions occur purely by chance. In the 'Moderate Rich's Society' model (Model A), where wealthy agents are risk-averse and less affluent agents are risk-seeking, the resulting wealth distribution is more balanced compared to the baseline model. Conversely, in the 'Rise of Meritocracy' model (Model B), where agents mistakenly believe they are in a meritocratic society, wealth inequality is exacerbated, with higher Gini coefficients and larger shares of wealth concentrated among the top 1% and 10% of agents.

These results underscore the need for policy formulation that takes into account the impact of agents' behavior on the overall distribution of wealth. Traditionally, income distribution policies have focused on direct intervention through taxation and redistribution, often overlooking the potential effects of changing agents' perceptions. Our findings suggest that a paradigm shift in policy-making is necessary.

However, it is crucial to acknowledge that this analysis is based on simulations, and we cannot definitively claim that these results represent a steady state. Moreover, the simplifications made for this thought experiment limit its direct applicability to real-world scenarios. To overcome these limitations, further research is needed that combines analytical methods and empirical data. By incorporating the latest findings from behavioral economics into agent mental models and using up-to-date data in market modeling, future studies can achieve more realistic and meaningful insights.

This study, which utilizes system dynamics and agent-based modeling, provides a robust framework for such future research. Enhancing the sophistication of these models with diverse behavioral insights and real-world data will enable the development of more effective and informed policies aimed at addressing economic inequality.

References

Cappelen, A. W., de Haan, T., & Tungodden, B. (2024). Fairness and limited information: Are people Bayesian meritocrats?. *Journal of Public Economics*, **233**, 105097.

Cruces, G., Perez-Truglia, R., & Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98, 100-112.

Drăgulescu, A., & Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. *Physica A: Statistical Mechanics and its Applications*, 299(1-2), 213-221.

Jones, C. I. (2015). Pareto and Piketty: The macroeconomics of top income and wealth inequality. *Journal of Economic Perspectives*, 29(1), 29-46.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449-1475.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1), 193-206.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

Lee, W. & Yoon, B. (2023). An Analysis of Impact of Income Transfer Policies on Wealth Distribution using Agent-Based Models(행위자기반모형을 활용한 소득보장정책이 부의 분포에 끼치는 영향 연구). *Korean Management Science Review(경영과학)*, 40(4), 67-80.

Lorscheid, I., Heine, B. O., & Meyer, M. (2012). Opening the 'black box' of simulations: Increased transparency and effective communication through the systematic design of experiments. *Computational and Mathematical Organization Theory*, 18(1), 22-62.

Pareto, V.(1897). “The New Theories of Economics,”, Journal of Political Economy, 5(4), 485-502.
Sandel, M. J. (2021). *The tyranny of merit: What's become of the common good?*. Penguin Books.

Silveira, G. A., Schrago, C. G., & Lopes, H. F. (2018). A Kolmogorov-Smirnov test for the molecular clock based on Bayesian ensembles of phylogenies. *PLOS ONE*, 13(1), e0190826.

Solt, F., Hu, Y., Hudson, K., Song, J., & Yu, D. E. (2016). Economic inequality and belief in meritocracy in the United States. *Research & Politics*, 3(4), 2053168016672101.

Sterman, J. D. (2001). System dynamics modeling: Tools for learning in a complex world. *California Management Review*, 43(4), 8-25.

Tao, Y., Wu, X., Zhou, T., Yan, W., Huang, Y., Yu, H., ... & Yakovenko, V. M. (2019). Exponential structure of income inequality: Evidence from 67 countries. *Journal of Computational Social Science*, 2(1), 3-30.

Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.