Philosophical Insights in Social Decision-Making: Integrating Trust, Loyalty, and Justice into Computational Models

Milad Kherghehandaz
Philosophy of science groups, Sharif University of Technology, Tehran, Iran

Abstract

Like game theory, classical decision-making models often fail to describe the complex social dimensions that influence human behavior. This paper, inspired by the theory of knowledge and existence from W. T. Stace, proposes a new approach to the decision-making model that integrates philosophical concepts—specifically trust, loyalty, and justice—into a mathematical framework for simulating social interactions. Inspired by Bayesian updating, reinforcement learning, and social belief propagation models, this approach adjusts trust and collaboration levels according to recognized social justice. using computational simulations, demonstrated that this model significantly enhances collaboration rates compared to classical models where self-interest dominates, like the Prisoner's Dilemma. Achieved results feature the critical role of justice recognition in promoting trust and loyalty, leading to constant social collaboration. This paper contributes to developing socially aware artificial intelligence, ethical decision-making frameworks, and policy design for real-world applications.

Keywords: "Bayesian Social Modeling", "Artificial Social Intelligence", "Social Decision-Making", "Trust Dynamics", "theory of knowledge"

Introduction

According to significant progress in artificial intelligence and social sciences, the need for more precise modeling of human behavior and social decision-making is becoming increasingly important [1]. One of the major challenges in this area is that existing models fail to fully simulate the complexities of human decisions, which are influenced by philosophical and social concepts [2]. Specifically, game models, such as the Prisoner's Dilemma, are fundamentally based on egotistic decision-making and optimizing instant rewards, which do not include complex factors, such as trust, loyalty, and justice, which play a Key role in social decision-making [3].

This research aims to develop a comprehensive and innovative model for simulating human social behavior that includes mathematical models, artificial intelligence algorithms, and philosophical and social concepts like trust, loyalty, and social justice in the decision-making processes of individuals in a society. Unlike classical models that only concentrate on instant rewards [4], this model considers more complicated social behaviors and can help simulate more accurate human decisions in various social and ethical contexts.

Walter Terence Stace (1993), in his work [5], explored concepts like categories and human cognitive processes. These philosophical concepts, particularly trust and loyalty, play a critical role in human social life and can have extreme effects on social interactions and individual decision-making. In this research, we specifically turn to these philosophical concepts and integrate them with Bayesian models to create a dynamic and modifying model for analyzing social behaviors.

In the next steps, first, we will review existing models and their limitations, then develop the philosophical and mathematical basis of the proposal, and finally, the proposed model will be compared with classical models. results of various simulations, particularly in analyzing social behaviors, trust, and loyalty, will be examined. This model has the potential to lead to wide applications in social artificial intelligence, recommendation systems, and social simulations in real-world scenarios.

Literature review

In recent decades, considerable efforts have been made to understand human social behavior and decision-making by computational models. However, most existing models, particularly classical game theory models like the Prisoner's Dilemma, have concentrated on the optimization of instant rewards and have largely ignored the underlying social and ethical dynamics that influence real-world decisions [6].

Game Theory Models and Decision-Making

Game theory models, like the Prisoner's Dilemma, have been generally utilized to study competitive behaviors. These models, which are useful for understanding the principles of collaboration and desertion, assume that individuals are self-interested and will make decisions based on instant rewards [7]. The simplicity of these models does not apprehend the complex social dynamics that occur in real-world situations, where factors such as trust, loyalty, and justice often play a key role in making decisions.
Furthermore, while game theory models have been utilitarian in understanding the rational behaviors of individuals, they often disappoint in simulating the moral and social influences that affect human decision-making. Harrison and McFadden (1974) and Bacharach (2006) have admitted the limitations of these models in capturing the complexity of human social behavior [8] [9].

Involve Social Concepts within Decision-Making

To address the limitations, researchers have surveyed the involvement of concepts such as trust, loyalty, and justice in computational models. Elster's (1989) and Fukuyama's (1995) studies on the role of trust in social capital emphasize the importance of these concepts in social decision-making. Trust, for instance, has been shown to impact collaboration and discord resolution, making it a key element in understanding human interactions [10] [11].
Also, loyalty and justice are necessary to fabricate long-term relationships and guarantee that individuals adhere to social norms and ethical standards. Studies by Deutsch (1973) on the psychology of collaboration proposes that when individuals recognize fairness in their social environment, they are more likely to collaborate, even despite instant actual costs [12].
In the social artificial intelligence (ASI) area, efforts have been made to model these social concepts. Researchers explore how AI systems can collaborate with human values and ethics standards in decision-making processes. For example, Wang, H et al. (2020) have worked on ethical decision-making algorithms that simulate moral reasoning in autonomous systems, while Yang, Q et al. (2020) have explored ways to make AI systems understand and respect human social dynamics [13] [14].

Philosophical Foundations of Social Behavior

The study of philosophical concepts through human cognition and social categories influences decision-making processes. W Stace's concept of moral and human knowledge supply insights into how abstract concepts like trust, loyalty, and justice can be modeled in AI systems. His study on the epistemology of knowledge suggests a framework for how human-like understanding can be simulated in computational systems [15].
Recently, there has been expanding interest in integrating philosophical models of cognition into AI [16]. Harnad (1990) has surveyed how cognitive processes, such as categorization and concept formation, can be modeled [17]. This theoretical approach corresponds to the research presented in this paper, which merges philosophical and mathematical models to simulate more accurate human-like decision-making.

Limitations of Existing Models

Despite the progress made in integrating social and philosophical concepts into decision-making models, many existing approaches still have trouble simulating the dynamics of social trust and moral reasoning comprehensively. Most models focus on the mathematical optimization of decisions without fully accounting for the developing nature of social trust and loyalty in real-world scenarios.

Methodology

Here, we present a methodology to develop a model that integrates concepts like trust, loyalty, and justice into the decision-making processes that individuals within a social context use. The model merges mathematical frameworks with philosophical insights to simulate human-like decision-making, especially in the context of social behavior.
The methodology includes three main components:
(1) Defining the Philosophical concepts
(2) Mathematical modeling
(3) Simulation of the model by Bayesian decision-making.

Defining the Philosophical Concepts

The foundation of the model is based on the philosophical concepts that influence human social behavior. These concepts are:
Trust (C): Trust is defined as the belief that another individual will act in a way that is beneficial or at least not harmful to oneself. In the model, trust is dynamic and can evolve based on the interactions between individuals. Trust can increase or decrease depending on the perceived actions of the other individual and the justice in the social environment.
Loyalty (L): Loyalty refers to an individual's commitment to the group or society, even when individual interests might conflict with group benefits. In the model, loyalty is influenced by the level of trust and social justice in the environment. As loyalty increases, individuals are more likely to cooperate, even when short-term rewards are minimal.
Justice (J): Justice is the perception of fairness in the distribution of rewards and resources within the society. In this model, individuals' actions are influenced by how fair they perceive the social environment to be. Higher levels of justice in the community lead to increased cooperation and social harmony.
These three concepts (trust, loyalty, and justice) are integrated into the decision-making process, allowing individuals to make more socially aware and ethical decisions.

Mathematical Modeling

We use a combination of Bayesian decision theory [18], propagation models [19], and a reinforced learning model [20] to model human decision-making. The primary elements of the mathematical model are as follows:
Trust Update Function: Trust is updated dynamically according to interactions with others. The trust value can be adjusted by paying attention to social justice and the behaviors observed in others. The trust function is represented as:

$$C_{i(t+1)} = C_{i(t)} + \alpha(C_{i(t)} - J_{soc}) \quad (1)$$

Which $C_{i(t)}$ is the trust of individual i at the time: t

j_{soc} is the attention justice in the society at time t

And α is a coefficient that defines how responsive the individual is considered to adjust in the social justice environment.
Loyalty Update Function: Loyalty is updated according to the individual's level of trust. The loyalty function is represented as follows:

$$L_{i(t+1)} = L_{i(t)} + \beta(L_{i(t)} - C_{i(t)}) \quad (2)$$

β is a coefficient that defines the responsiveness of loyalty corresponding to trust.

Action Decision: The decision to collaborate or fail is influenced by trust, loyalty, and justice. The decision-making function is represented probabilistically, where individuals decide to act based on the level of loyalty and trust in the society and the recognized justice. The probability of collaboration is given by:

$$P_{act(i)} = \frac{L_{i(t)}.C_{i(t)}}{1+J_{soc}} \quad (3)$$

$P_{act(i)}$ is the probability that individual i will cooperate (see Appendix 1).

Simulation of the Model

The model is performed and simulated by Bayesian decision-making and fuzzy logic to account for uncertainties and ambiguities in human decision-making. The simulation process includes the following steps:
Initial Setup: A population of agents (individuals) is initialized with random values for trust, loyalty, and justice. These individuals interact with each other according to rules introduced for updating trust and loyalty.
Iterative Decision-Making: Over several cycles, each interacts with others, updating their trust and loyalty based on the examined actions of others. The decision to collaborate or fail is made probabilistically based on the current values of trust, loyalty, and justice recognized from the environment.
Updating Social Justice: Social justice is updated after each round based on the collective behavior of the individuals. If a large amount of the population acts collaboratively, the recognized justice in the society increases, which in turn encourages further collaboration.
Trace Performance: The performance of the model is traced by examining the overall collaboration rate in the society and the average loyalty and trust values across individuals. Also, it has been analyzed that the system's how response to different levels of social justice.

Performance and Algorithm

The algorithm of simulation is as follows:
Determine a population of N individuals with random primary values for trust, loyalty, and justice.
Update trust and loyalty based on interactions with others for each one.
Calculate the probability of collaboration using the decision function.
Update social justice based on the collective behavior of the population.
Trace the collaboration rate, trust, and loyalty for each over time.
Repeat steps 2-5 for T cycle.

Computational Tools

The model is implemented using Python, and the following libraries are used:
NumPy for numerical computations.
Matplotlib for visualizing the results.
SciPy for statistical analysis and simulations.
The simulation allows for a flexible and scalable framework for modeling social decision-making in various environments.

Results

Here, we present the results of the simulations. The main impartial of these simulations is to illustrate how the integration concepts like trust, loyalty, and justice can improve the simulation of social decision-making compared to classical models, such as the Prisoner's Dilemma. The results emphasize the impact of these concepts on individual behaviors, the overall collaboration rate in society, and how different levels of social justice influence the decision-making of each individual.

Experimental Setup

We used a population of N=100 individuals, each with random primary values for trust, loyalty, and justice. The individuals interact with each other over T=50 cycles, during which their trust and loyalty are updated according to interactions with others. The probability of collaboration is calculated using the dynamic decision function that integrates trust, loyalty, and justice. Then, Social justice is updated after each cycle according to the collective behavior of the population.
The simulations are run with varying levels of initial social justice, which can be the standard level of fairness in the society. The results are traced by examining the collaboration rate, trust, and loyalty over time.

Simulation results

a) Impact of Trust and Loyalty on Collaboration

One of the main results of the simulations is the important impact of trust and loyalty on the likelihood of collaboration. As the trust between individuals increases, the overall collaboration rate in the society also increases. This result is in contrast to the classical Prisoner's Dilemma, where individuals often fail, driven only by instant rewards. In this model, higher levels of trust lead to a more collaborative society, even lacking short-term rewards.
Figure 1 shows the relationship between the average trust and collaboration rate. As trust increases, the collaboration rate ambit higher values, indicating that individuals are more likely to collaborate when they trust each other.
b) Social Justice's Effect on Decision-Making
The results also illustrate that social justice plays a crucial role in making individual behaviors. When individuals recognize the society's fairness and justice, they are more likely to collaborate, even if collaboration results in a personal cost, which is compatible with findings from social psychology (e.g. Deutsch, 1973). This suggests that individuals are more likely to collaborate in a fair environment.
Figure 2 illustrates how different levels of social justice impact the long-term collaboration rate. In societies with higher recognized justice, individuals are more disposed to collaborate.
c) Comparison with Classical Models
To compare the achieved results with classical models, we simulated the Prisoner's Dilemma using the same population size and number of cycles but without the insertion of trust, loyalty, or justice.
Figure 3 compares the collaboration rate in the classical Prisoner's Dilemma and our model. In the classical game theory, the collaboration rate remains low due to the self-interested nature of the individuals. In contrast, our model shows a significantly higher collaboration rate, especially when trust and justice attending.
d) Changes in Trust and Loyalty
Also, in addition to the overall collaboration rate, we analyzed the dynamic changes in trust and loyalty over time. At first, the simulation starts with random initial values, but over time, trust and loyalty grow according to interactions with others.
Figure 4 shows the change in trust and loyalty over time. As individuals interact and examine each other's actions, trust and loyalty are updated dynamically. This leads to the emergence of collaborative behaviors over time, further enhancing social cooperation.

Key Observations and Insights
Trust and Loyalty Lead to Collaboration: As individuals trust each other more, they are more likely to collaborate, which finally leads to an increase in the overall collaboration rate in society.
Social Justice motivates collaboration: Societies with higher levels of recognized justice motivate individuals to act more collaboratively. This considers the importance of fairness in social interactions.
Dynamic Trust and Loyalty: As trust and loyalty are not fixed, they grow over time so that individuals interact and examine each other's actions. This dynamic nature of trust and loyalty is crucial in apprehending the complexities of real-world decision-making.
upgrade relation to Classical Models: The addition of philosophical concepts such as trust and justice significantly improves the ability to simulate social decision-making and the evolution of collaboration in societies, compared to classical models like the Prisoner's Dilemma, where collaboration is ordinarily low.

Graphic display of Results
The following figures illustrate the key findings of the simulations:

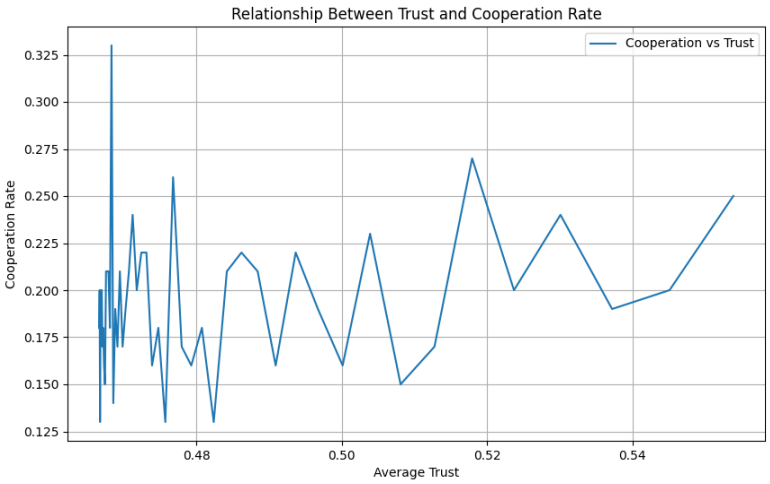

Figure 1: Relationship between trust and collaboration rate.

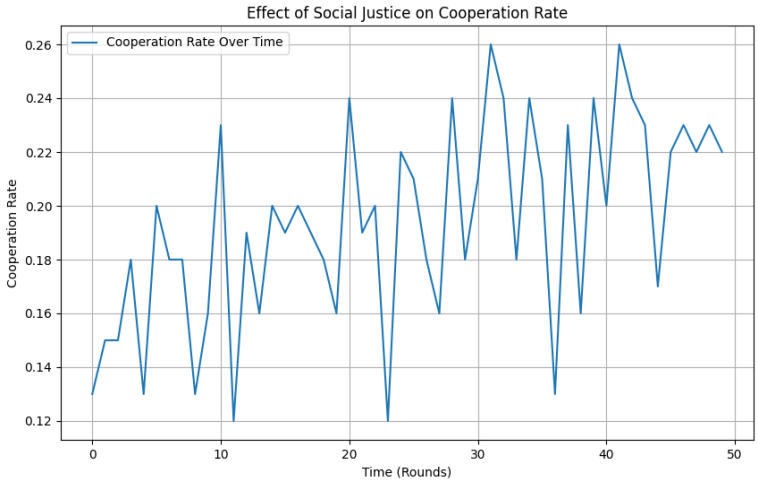

Figure 2: Impact of social justice on the overall collaboration rate.

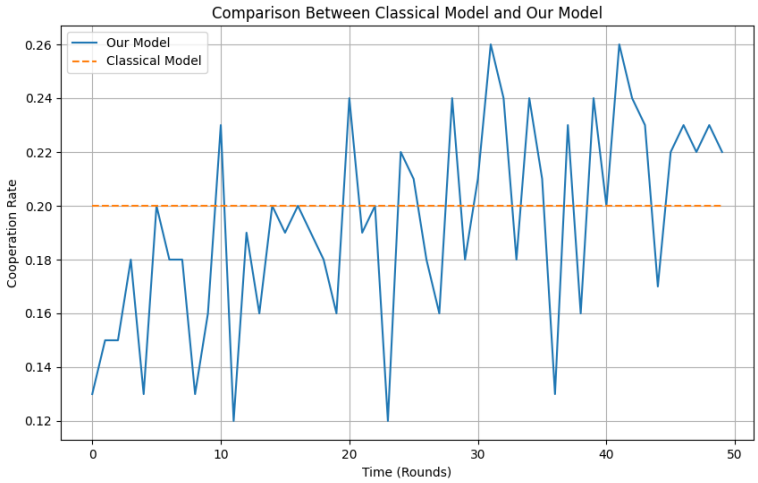

Figure 3: Comparison of collaboration rate between the classical Prisoner's Dilemma and this model.

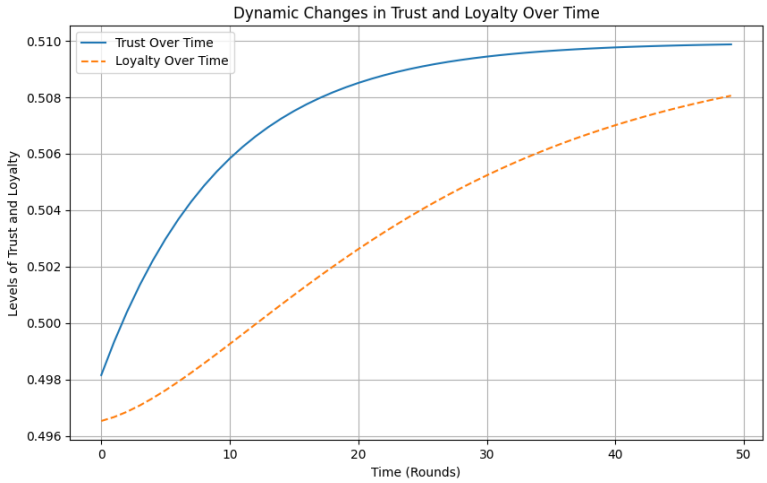

Figure 4: Trust and loyalty Dynamic changes over time.

Discussion

In this section, the results of the simulations presented earlier have been analyzed. Here, we discuss the implications of the trust, loyalty, and justice factors on human social behavior and decision-making. The first concentrates on the insights achieved from the model and how the integration of philosophical concepts significantly improves the simulation of social behavior compared to classical models.

Trust and Loyalty on Collaboration Effects

The crucial role of trust and loyalty is one of the important results achieved from simulations. In classical models such as the Prisoner's Dilemma, collaboration is disposed low as individuals pursue maximizing their instant costs. However, this model, which integrates trust and loyalty, shows that as trust increases between individuals, the likelihood of collaboration rises remarkably. This is consistent with social psychology theories, like those proposed by Deutsch (1973), which suggest that trust is a fundamental driver of collaborative behavior. While loyalty is often seen as a long-term commitment, in this model, it results in more collaborative behaviors, even in situations where instant personal costs are minimal. This is a significant change from classical models, where loyalty is often not considered, and decision-making is mainly based on maximizing short-term costs.

Social Justice as a Stimulus for Collaboration

The addition of social justice in this model supplied another constraining result: individuals are far more likely to collaborate when they recognize the society they belong to as just. In societies where justice is recognized to be high, individuals tend to exhibit higher levels of collaboration, even when personal sacrifice is required. This result strengthens insights from Fukuyama (1995) and Elster (1989), who argue that a sense of justice is intrinsic to the functioning of social systems and motivates individuals to contribute to collaborative actions. This model also shows that justice has a cascading effect: when individuals act collaboratively in a sense of justice, it fosters more justice and finally creates a positive feedback loop that motivates further collaboration.

Changes in Trust and Loyalty dynamically

The dynamic nature of trust and loyalty is another important attention from the simulations. Contrary to classical models, which suppose stable trust values, this model considers the growing nature of human relationships. Trust and loyalty oscillate according to social interactions, decisions, and the recognition of fairness in society. In addition, Individuals who practiced constructive interactions with others tended to increase their trust and loyalty, which finally caused higher rates of collaboration. This aspect of the model makes it more pragmatic and adjustable to the changing social environment. Also, dynamic nature is an advancement in simulating real-world social systems, where relationships are regularly evolving and affected by individual interactions.

Applications in Policy-Making and Compromise

These results could be applied to real-world scenarios in subjects like policy-making, compromise, and negotiation. Understanding how trust, loyalty, and justice affect group decision-making could help policymakers create more effective strategies for strengthening collaboration in diverse social and political environments.

Integrating Environmental and Circumstantial Factors

While this model concentrates on individual interactions and abstract concepts, future work could integrate environmental factors such as economic conditions, cultural influences, and geopolitical dynamics that make decision-making in different circumstances.

Exploring Long-Term Results

Finally, future research could probe the long-term effects of trust, loyalty, and justice on public stability and individual well-being. While the current model captures short-term cooperation, understanding the long-term consequences of these social concepts on social cohesion and sustainable cooperation could provide further insights.

Conclusion

Modeling social behavior by integrating philosophical concepts into decision-making frameworks has been proposed in this study. Here, we tried to introduce a new approach to do it more precisely. The results show that integrating trust, loyalty, and justice significantly improves the accuracy and realism of social decision-making models. this results also show higher collaboration rates and a more broad understanding of human behavior in social environment. due to expanding the model to include diverse recognition of justice, group dynamics, and more complex ethical reasoning, further advancing the field of social artificial intelligence and its applications.

References

[1] Y. E. J. a. D. Y. Duan, " Artificial intelligence for decision making in the era of Big Data–evolution, challenges

and research agenda," *International journal of information management, 48,* pp. 63-71, 2019.

[2] P. Glynn, " Integrated Environmental Modelling: human decisions, human challenges," *Special Publications, 408(1),* pp. 161-182, 2017.

[3] E. Barron, Game theory: an introduction, John Wiley & Sons, 2024.

[4] S. R. P. B. R. a. R. K. Samudrala, "A Game Theoretic Cognitive Spectrum Sensing Scheme for IoT Networks," Telecommunications and Radio Engineering, 83(9), 2024. [Online].

[5] W. Stace, The theory of knowledge and existence, 1993.

[6] M. C. C. C. D. D. C. R. D. C. D. R. E. B. G. J. G. N. H. C. a. H. D. Calder, "Computational modelling for decision-making: where, why, what, who and how," *Royal Society open science, 5(6), p.172096.,* 2018.

[7] R. Axelrod, "the evolution of cooperation," *Journal of theoretical biology, 299,* pp. 21-24, 2012.

[8] D. McFadden, "The measurement of urban travel demand.," *Journal of public economics, 3(4),* pp. 303-328, 1974.

[9] M. Bacharach, Beyond individual choice: teams and frames in game theory, Princeton University Press, 2006.

[10] J. Elster, Solomonic judgements: Studies in the limitation of rationality, Cambridge University Press, 1989.

[11] F. Fukuyama, "Social capital and the global economy," Foreign Aff., 74, 1995.

[12] M. Deutsch, "A theory of cooperation–competition and beyond.," in *Handbook of theories of social psychology 2*, 2012, pp. 275-294.

[13] H. K. A. C. D. a. L. T. Wang, "Ethical decision making in autonomous vehicles: Challenges and research progress," *IEEE Intelligent Transportation Systems Magazine, 14(1),* pp. 6-17, 2020.

[14] Q. S. A. R. C. a. Z. J. Yang, "Re-examining whether, why, and how human-AI interaction is uniquely difficult to design," in *In Proceedings of the 2020 chi conference on human factors in computing systems*, 2020.

[15] W. Stace, The concept of morals., 1937.

[16] Y. Maruyama, Symbolic and statistical theories of cognition: towards integrated artificial intelligence In Software Engineering and Formal Methods. SEFM 2020 Collocated Workshops: ASYDE, CIFMA, and CoSim-CPS, Amsterdam, The Netherlands, September 14–15, 2020, Revised, Springer International Publishing, 2021.

[17] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena, 42(1-3),* pp. 335-346, 1990.

[18] A. a. R. R. Wagner, "Inhibition in Pavlovian conditioning: Application of a theory," *Inhibition and learning,* pp. 301-336, 1972.

[19] R. a. B. A. Sutton, "Reinforcement learning," *Journal of Cognitive Neuroscience, 11(1),* pp. 126-134, 1999.

[20] B. a. J. M. Golub, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics, 2(1),* pp. 112-149, 2010.

Appendix: Derivation of Equations used in model

Equations can be derived based on the following principles:

1. Bayesian Updating in Social Learning

In Bayesian learning models, an individual updates their belief about a variable (θ) by integrating new information (X), adjusting their prior estimate:

$$P(\theta \mid X) = \frac{P(X \mid \theta)P(\theta)}{P(X)}$$

2. Reinforcement Learning and Weighted Belief Adjustment

In reinforcement learning models, updates follow a similar principle: beliefs (or policies) are adjusted according to a prediction error. The general update equation is:

$$V_{s(t+1)} = V_{s(t)} + \alpha(V_{s(t)} - R_t)$$

where:

$V_{s(t)}$ is the estimated value at time t

R_t is the recognized reward,

α is the learning rate.

3. Belief Propagation in Social Networks

Also, in belief propagation models used in social network theory, individuals modify their beliefs according to interactions with their environment. The rate of trust adjustment is governed by α, which determines whether individuals adjust their trust gradually (low α) or rapidly (high α).

To model the dynamic nature of trust C_i We employ an adaptive update mechanism inspired by theories of Bayesian updating, reinforcement learning, and belief propagation in social networks. Trust is not a static variable; rather, it evolves based on individual experiences and perceived social justice J_{soc}. adaptive learning principle follow as:

$$X_{i(t+1)} = X_{i(t)} + \gamma(X_{i(t)} - Y_{(const)})$$

where:

$X_{i(t)}$ is any evolving variable in the model (trust, loyalty, cooperation tendency, or another social factor).

$Y_{(const)}$ represents the contextual factor influencing $X_{i(t)}$, such as perceived social justice, peer cooperation levels, or cultural influences.

γ is a learning rate coefficient that determines the responsiveness of the variable to contextual changes. This structure can be extended to other social and cognitive variables. Our model follows the same principle: $C_{i(t+1)} = C_{i(t)} + \alpha(C_{i(t)} - J_{soc})$

Trust is updated in response to the perceived fairness of the environment. If an individual experiences fairness $C_{i(t)} > J_{soc}$ their trust increases. If they perceive injustice $C_{i(t)} < J_{soc}$ their trust declines.

The adjustive update mechanism used for trust $C_{i(t)}$ can be generalized to other variables in the model, like loyalty $L_{i(t)}$ and collaboration probability $P_{i(act)}$. This generalization allows the model to apprehend dynamic changes in social behaviors according to individual recognized social conditions. By applying this to loyalty, we have:

$$L_{i(t+1)} = L_{i(t)} + \beta(L_{i(t)} - C_{i(t)})$$

where:

Loyalty is adjusted according to an individual's level of trust.

β represents the justification rate, determining how rapidly loyalty responds to trust changes. This equation guarantees that loyalty gradually increases when trust is elevated and decreases when trust is declined, reflecting real-world social interactions.

Also, the decision to collaborate can be modeled using a probabilistic function according to trust, loyalty, and justice:

$$P_{act(i)} = \frac{L_{i(t)}.C_{i(t)}}{1 + J_{soc}}$$

where:

The probability of collaboration increases when trust and loyalty are elevated.

Elevator social justice leads to a more collaborative society.

The divisor guarantees restricted probabilities while maintaining dynamic justification. We can see that each equation reflects a learning-driven adjustment, guaranteeing that the model remains adaptive to social and cognitive changes rather than being static.

Justifying the Equations in Our Model follow as:

- The equations guaranteeing that trust grow dynamically rather than remaining static, making it more realistic in simulating social behaviors.
- It apprehends gradual adaptation rather than rapid change, corresponding to observed human decision-making processes.
- It provides a mathematically interpretable formulation for how individuals adjust their trust in response to recognized justice.

Thus, this trust update function extends classical models of Bayesian belief modification, reinforcement learning, and social network dynamics to create a more realistic representation of human decision-making in a social environment.