The Dynamics of Cooperation with Commitment in A Population of Heterogeneous Preferences–An ABM Study

Wei Wang (weiwang@bupt.edu.cn)
Luzhan Yuan (johnyuan@bupt.edu.cn)
Zheng Jiang (jiangzheng@bupt.edu.cn)
Gaowei Zhang (zhanggaowei@bupt.edu.cn)
Yi Wang (yiwang@bupt.edu.cn)
Beijing University of Posts and Telecommunications

Abstract

Prior literature shows that some mechanisms, e.g., commitment, could give rise to cooperation. However, participants' diverse propensities to cooperate may limit such mechanisms' effectiveness. Thus, we bring individual differences in their propensities to cooperate into the reasoning of long-term social dynamics of cooperation through an agent-based modeling approach. Our results suggest that commitment may still guarantee cooperation when individuals have different propensities to cooperate but has weaker effects, and the setups of commitment are also important. Our study highlights the importance of integrating individual preferences in analyzing collective dynamics of a population consisting of individuals of heterogeneous characteristics, thus offering implications to facilitate cooperation in rich real-world scenarios.

Keywords: commitment; cooperation; agent-based modeling

Introduction

Interpersonal cooperation is a fundamental component of many modern social and economic activities, however, it neither emerges from nothing nor maintains and diffuses automatically (Zheng, Veinott, Bos, Olson, & Olson, 2002; Wang & Redmiles, 2016b; Moisan, ten Brincke, Murphy, & Gonzalez, 2018). Cooperation is not an arbitrary choice but could be deliberated strategic behaviors in social productions (Wang & Redmiles, 2016a). When individual members decide to maximize their own short-term benefits independently, *social dilemma*, in which socially optimal could never be achieved, can happen (Olson Jr, 1971). Nobel Laureate Elinor Ostrom pointed out, in the collective actions driving social production, if many choose to be uncooperative, cooperators' risk would significantly increase and become more doubtful of continuing to be cooperative (Ostrom, 2003), eventually resulting in a *tragedy of commons* where no one cooperates.

Fortunately, human beings have found various mechanisms, such as incentives, social norms, etc., to resolve social dilemmas and develop, maintain, and enforce cooperation (Axelrod & Hamilton, 1981; Nowak, 2006; Hauser, Hilbe, Chatterjee, & Nowak, 2019). Commitment is one such mechanism. Literature in multiple disciplines has confirmed commitment's positive effects on solving social dilemmas and guaranteeing cooperation from both theoretical and empirical perspectives (Back & Flache, 2006; Chen & Komorita, 1994; Corbett & Le Dantec, 2018; Han, Pereira, & Santos, 2012; Sasaki, Okada, Uchida, & Chen, 2015; Pearce, Branyiczki, & Bigley, 2000). However, most extant literature on commitment mechanisms assumes that there is a homogeneous population of members whose individual preferences were largely neglected. In fact, such an assumption may not hold. Individual differences in propensity to cooperate are consistently reported to have significant impacts on their behaviors (Kortenkamp & Moore, 2006; Bowles & Gintis, 2004; Thielmann, Spadaro, & Balliet, 2020). Besides, modern work organizations are increasingly globally distributed, different people's backgrounds and the lack of collocated context could lead to more diverse individual preferences in cooperation (Morrison-Smith & Ruiz, 2020; Ajmeri, Guo, Murukannaiah, & Singh, 2020), amplifying such differences' impacts on cooperation with the commitment mechanism. Moreover, as Maxwell & Oliver pointed out in their influential book (1993), even a small proportion of people with heterogeneous characteristics in a group, e.g., highly-motivated, could lead to outcomes which would be impossible to obtain with groups with strictly homogeneous characteristics.

Therefore, there is an imperative to develop an in-depth understanding of how diverse individual preferences influence the effects of commitment on cooperation. We thus have the first research questions.

RQ$_1$: *Do diverse preferences influence the development of cooperation with the commitment mechanism?*

If the diverse individual preferences' effects could be confirmed, since prior literature (Han et al., 2012) shows that the effects of commitment largely determined by its setups, we would like to examine how these setups work when introducing individual preferences in cooperation; thus, we have the second research question:

RQ$_2$: *How do different setups of the commitment influence the cooperation of people of diverse preferences?*

This article reports our efforts in answering the above research questions through the agent-based modeling (ABM) technique (Ren & Kraut, 2014; Bonabeau, 2002). Leverage game theory, we built a model to simulate how individuals (*agents*) with different preferences in cooperation make strategic decisions in interacting with other members in a fixed population, with *commitment* as a type of strategies in dyadic interactions yet publicly-visible for all members. We then allowed the agents to interact with each other in a discrete event simulation setting and experimented with a wide range of model parameters, which enabled us to explore the long-term dynamics of cooperation.

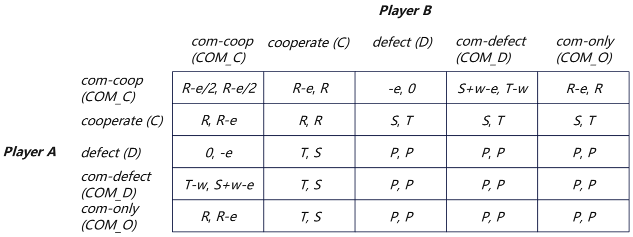
In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

The ABM Model

Theoretical Foundations

A Game Extending Social Dilemma with Commitment

To incorporate commitment into social dilemmas, we extended the classic social dilemma. Since proposing commitment is an action, it could be viewed as a part of strategies. A commitment, once made, may not always be fulfilled later. Thus, we need two extra strategies. One is for making a commitment and fulfilling it, and another is for making but not fulfilling. In addition, some members may wait for other parties' commitment as the prerequisite of their own cooperative behaviors; we need another extra strategy. Therefore, there are potentially five strategies. That is similar to the model in Han (2016). We slightly revise their game as follows:



		Player B				
		com-coop (COM_C)	cooperate (C)	defect (D)	com-defect (COM_D)	com-only (COM_O)
	com-coop (COM_C)	R-e/2, R-e/2	R-e, R	-e, 0	S+w-e, T-w	R-e, R
	cooperate (C)	R, R-e	R, R	S, T	S, T	S, T
Player A	defect (D)	0, -e	T, S	P, P	P, P	P, P
	com-defect (COM_D)	T-w, S+w-e	T, S	P, P	P, P	P, P
	com-only (COM_O)	R, R-e	T, S	P, P	P, P	P, P

Figure 1: The adapted game with commitment mechanisms.

Compared with the classic prisoner's dilemma, the extended game (Fig. 1) has three new strategies, *COM_C*, *COM_D*, and *COM_O*. *COM_C* refers to the action of making a commitment first and then fulfilling it with cooperative behaviors. *COM_D* makes a commitment but never fulfills it, which is more or less antisocial. *COM_O* means one will never cooperate unless it involves a commitment. This strategy differs slightly from the others. While conventional prisoner's dilemma assumes agents behave simultaneously, *COM_O* allows some implicit sequential decisions. Here, agents using *COM_O* may wait a short period to observe the other party's action, then decide their action accordingly. We do not break it into two separate actions because observation itself does not yield any changes to payoff.

The introduction of the new strategies also leads to some changes to the payoffs, particularly when one uses *COM_C* in interaction. First, when both players use *COM_C*, they would share the cost of commitment so each receives $R - \frac{e}{2}$. When the other plays *C* or *COM_O*, the *COM_C* player would bear all cost. When meeting a *D* player, the *COM_C* player has no payoff but still needed to bear the cost. However, if the other party enter the commitment but never honor it (*COM_D*), might be charged some penalty *w* in some form, e.g., losing reputation, while the *COM_C* player receives *S-e+w*. Thus, the extended game structure allows us to describe and analyze interactions among agents in social production. Agents could use the payoff structure to evaluate their expected payoff and decide which strategy to use.

Propensity to Cooperate In a prisoner's dilemma, a rational player's strategy would be *defect*. But in the real-world, "human beings can be anything but rational..," as Lester Lave commented (1962). Some people always cooperate, while some always defect. To cooperate or not, is never a simple rational choice based on immediate gains (Young, 1998).

We thus need to model the distributions of people's different propensities to cooperate in the population. We extracted it based on real-world prisoner's dilemma experiments. Jones (Jones, 2008) surveyed 36 studies of real-world experiments of prisoner's dilemma from 1959 to 2003 and found that the median rate of cooperation is 39%, with a 19% minimum and an 80% maximum. Jones' results suggest that people strongly prefer to *cooperate* or strongly prefer to *defect* would be less likely to lower than 20%. Recent studies in the lab and natural settings (Balliet, Li, Macfarlan, & Van Vugt, 2011) also suggest similar results. So, when we model individual preferences, we would bring these insights into the model by allowing the initial distributions of individual strategies to fall into proper intervals. Furthermore, we would also experiment with different parameters of the distributions of agents' initial strategies. In our work, we break an individual's payoff into two parts: the idiosyncratic payoff resulting from satisfying individual preferences and the interaction payoff received from interacting with another agent.

Agents: How They Make Decisions and Behaves?

Fig. 2 describes a simplified interaction scenario between two agents. In this scenario, agents *A* and *B* are chosen to interact with each other. We describe the decision-making process from the agent *A*'s perspective. For the agent *B*, the decision-making process is identical (see Fig. 2), because they are opponents to each other in a symmetrical game.

In this process, agents first retrieve their opponents' historical interaction information and use it as the basis for their decision-making (Simon, Fagley, & Halleran, 2004). First, the **bounded rationality** limits agents' capability in retrieving and processing information. Thus, only a part of the history of one's opponent's behaviors, rather than the entire history, could be used in the decision-making. Here, we allow an agent to only use the opponent's latest *m* interactions as the references for the decision-making.

Now, we are going to incorporate different individual preferences. The total payoffs include idiosyncratic payoff resulting from one's individual preference and interaction payoff ($P_{interaction}(s, \_s)$ in Equ. 1) resulting from interacting with another agent under the game specified in Fig. 1. Doing so allows us to express the individual's **propensity to cooperate** as a part of an agent's utility. Supposing an agent *A* uses strategy *s*, the opponent uses *_s*, the agent's payoff is:

$$U_A(s, \_s) = f_A(s) + P_{interaction}(s, \_s) \qquad (1)$$

where $f_A(s)$ is a personalized function defined over the five strategies to represent the impact of an individual's preference in cooperation on one's utility. It may take any form, but preserving discrete partial orders.

When the agent *A* plays strategy *s*, agent *B*'s counter strategy *_s* could be one of the five strategies. Since the history

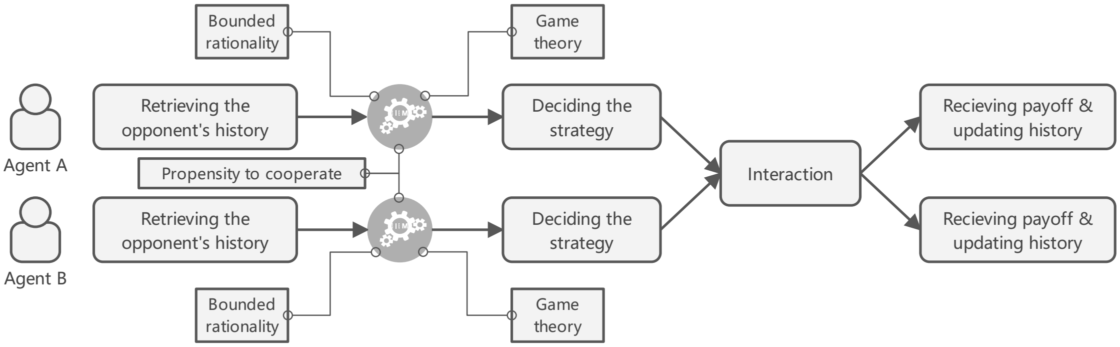


Figure 2: A simplified interaction scenario between two agents featuring their decision-making process.

(of the m size) gives a possible distribution of $\_s$ over all the five strategies, let us use k_1 to k_5 to represent the frequency of a specific strategy in agent B's history. An agent's expected payoff of using strategy s can be written as:

$$EP_A(s,\_s) = \frac{k_1}{m} \times U_i(s, COM\_C) + \frac{k_2}{m} \times U_i(s, C) + \frac{k_3}{m} \times U_i(s, D) + \frac{k_4}{m} \times U_i(s, COM\_D) + \frac{k_5}{m} \times U_i(s, COM\_O) \quad (2)$$

Given that we assume our agents are reasonably rational, they would choose the strategy *s'* that *maximizes* the expected payoff in the interaction with agent B. Then, the agent would behave accordingly to turn the decision into behavior. The agent B's decision-making is identical, but with the agent A's history as the reference. Obviously, their decisions are essentially the "*best-replies*" to each other (Young, 1998; Gibbons & Gibbons, 1992; Fudenberg & Tirole, 1991). However, in the real world, people often cannot calculate the expected payoffs precisely as we do above. They may only have some ambiguous guesses about the payoffs. Therefore, their decisions are often not deterministic. I.e., they may believe that it is possible that strategy X yields a better payoff than Y does, but they are not 100% sure about that. To describe the impact of ambiguity in decision-making, we follow the conventions and use the logistic learning rule (Fudenberg & Levine, 1998) to specify the probability of choosing strategies. For a specific strategy s_i $(1 \leq i \leq 5)$, the probability for the agent A to use it is:

$$\frac{e^{EP_A(s_i,\_s)}}{\sum_{i=1}^{5} e^{EP_A(s_i,\_s)}} \quad (3)$$

Thus, our agent could accommodate the real-world uncertainty in decision-making. In this setting, if a strategy is likely to yield a higher payoff, it would be more likely but not definitely to be chosen as the strategy in the coming interaction. To turn the probabilities into a deterministic strategy in an interaction, the ABM utilizes the urn randomization technique (Ross, 2014) to select the strategy. However, in some cases, agent A may not be that calculated, i.e., the bounded rationality may lead to mistakes in decision-making at a small but non-zero probability. We use θ to denote the probability of making mistakes.

Collective Dynamics

Developing cooperation might require multiple rounds of interactions. Therefore, our ABM model shall accommodate discrete-event feature for the purpose of analyzing long-term dynamics. During each round, two agents are randomly selected to interact. They follow the decision-making process described above to make decisions and interact. By continuing simulating interactions, the collective dynamics at the community level can automatically emerge and analyzed.

Model Implementation

The model is implemented with PYTHON's *Mesa* (Kazil, Masad, & Crooks, 2020) ABM framework. An agent is implemented as a class that encapsulates agents' personal propensity to cooperate, decision-making, and behavioral history. We use an independent process to control and monitors the entire simulation process, including initializing the simulation, selecting participating agents in each period, tracking the periods, and monitoring and logging the entire simulation during the predefined number of periods (1,000 in our study, see Tab. 1. The ABM is thus highly modifiable and extensible to incorporating other social factors and can be adapted as a workbench for other studies. The source code of the ABM model is publicly available at: `https://figshare.com/articles/software/code/24138345`.

Simulation Experiment Design

With the above ABM model, we thereby design two simulation experiments to answer the $\mathbf{RQ_1}$ and $\mathbf{RQ_2}$ accordingly.

Experiment 1 for RQ_1

Since we want to check if individuals' diverse preferences have any effects, we shall compare two conditions with and without considering individual preferences. Thus, simulations without individual preference could be considered as the "*control*" group, while simulations with individual preferences serve the "*treatment* group. The only difference between the two groups is the presence of individual preference in cooperation.

Experiment Process Fig. 3 describes the design of the experiment. Simulations for both conditions are basically the

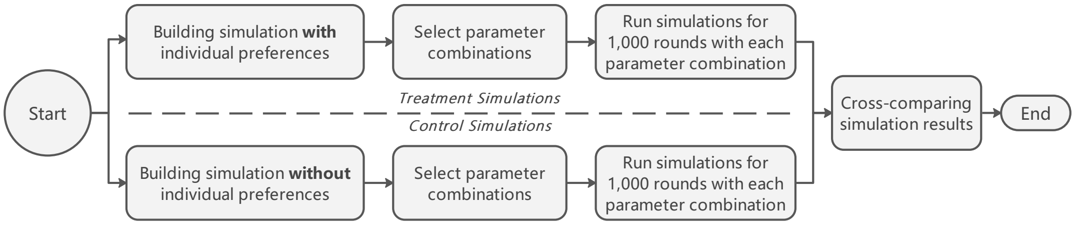

Start
Building simulation **with** individual preferences
Select parameter combinations
Run simulations for 1,000 rounds with each parameter combination
Treatment Simulations
Control Simulations
Building simulation **without** individual preferences
Select parameter combinations
Run simulations for 1,000 rounds with each parameter combination
Cross-comparing simulation results
End

Figure 3: The overall process of the experiment 1.

same, except that the treatment uses the payoff structure consisting of both idiosyncratic and interaction parts, while the control only has the interaction part. Then, we select simulation parameters based on prior literature and empirical evidence. Of course, in the treatment condition, we also need to determine the extra parameters related to individual preference in addition to parameter combinations in the control condition. After this step, we run 1,000 rounds of simulations for each parameter combination under two conditions. Once all simulations are concluded, we cross-compare the results to check if there are significant differences between the control and treatment conditions.

Experiment Parameters In this experiment and the second, parameters are determined following several simple heuristics. First, parameters should satisfy the quantitative relationships and other necessary conditions specified in related theories. E.g., a social dilemma requires $T > R > P > S$. Some specific relationships identify some critical boundary conditions. We reuse such boundary conditions to guide our experiments. Second, we leverage the rich empirical literature to set reasonable parameters. For example, the initial composition of agents could be determined by the evidence from many empirical prisoner's dilemma experiments and their meta-analysis (Jones, 2008). Third, setting the parameters should also follow the common sense. For example, it is might not proper to allow magnitudinal differences between one's idiosyncratic payoff and interaction payoff. Finally, to avoid excessive computations due to trivial numerical differences, we set the minimal difference in a quantitative relationship as 0.1. Bearing the above in mind, we set the following default values to the parameters (Tab. 1). Besides, the population size is fixed as 50 agents.

Experiment 2 for RQ$_2$

Prior literature shows that commitment is an effective mechanism to carry social dilemmas to cooperation under certain setups of the commitment (Han et al., 2012; Han, 2016). These setups are often specified in the quantitative relationships between the cost of commitment (e) and the penalty (w) to those who commit but not deliver. One such theoretical condition is given in Han et al. (Han et al., 2012), which claims that when $e < \frac{2R}{5}$, and $w > max\{\frac{T-R-S}{2} - \frac{3e}{4}, \frac{T-R-2S}{3} + \frac{5e}{6}\}$, the cooperation is more likely to be achieved. The second experiment would use such theoretical results to guide our experiment.

With the default parameters, the condition in Han et al. is $e < 0.24$, and $w > max\{0.2 - \frac{3e}{4}, 0.13 + \frac{5e}{6}\}$. If plotting them in a coordinate, the condition defines an area marked with dashes in the background of Fig. 5. This area is mostly in around the upper left corner where the cost of commitment is small but the penalty is large. According to Han et al., (Han et al., 2012), when $< e, w >$ falls into this area, it is would more probable to have a large proportion of individual agents use strategies leaning to cooperate (C and COM_C). If such a condition still holds with diverse individual propensity to cooperate, we could use it in the design of commitment mechanisms. Experiment 2 is designed to verify this.

Experiment 2 does not need the control and treatment conditions. Instead, we let the parameters e and w vary in the range [0.1, 0.5]. Since the minimal step is 0.1, both of them could be 0.1, 0.2, 0.3, 0.4, and 0.5, making that there are 25 combinations of $< e, w >$ in total from $< 0.1, 0.1 >$ to $< 0.5, 0.5$. We then run 100 rounds of simulations with all these 25 combinations. For each combination, we compute the average proportion of agents using strategies leaning to cooperation when their behaviors become stable. By comparing such proportions under the 25 combinations, we can check if the theoretical results still hold when considering individual preferences.

Results & Findings

This section reports on the results of our study. Note that all simulations have been running multiple rounds in our experiments, so the results are aggregated averages of all simulation rounds rather than a single run of each simulation to avoid random errors resulting from a single simulation run (Zeigler, Praehofer, & Kim, 2000).

Experiment 1 Results & Findings

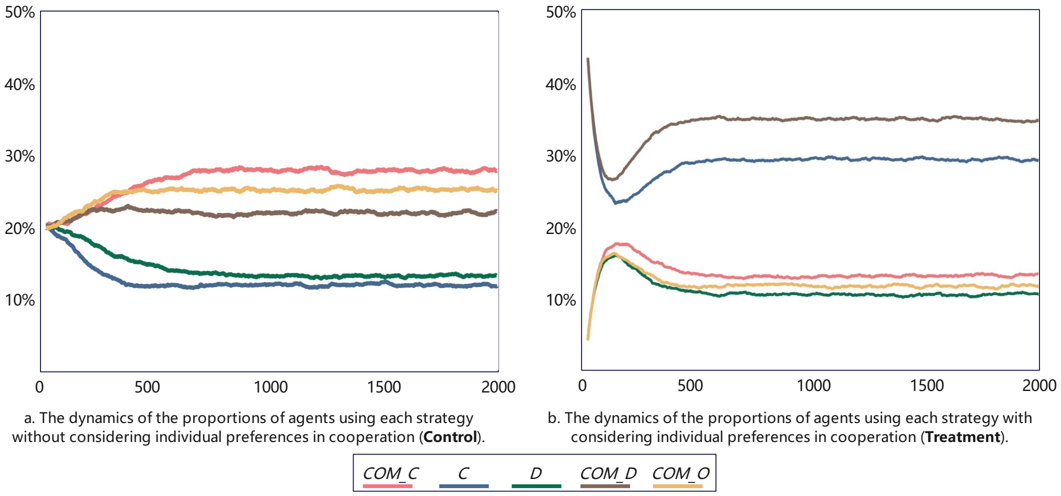
Experiment 1 attempts to compare the dynamics of cooperation under two conditions: with and without considering individual preferences in cooperation. Fig. 4 describes the aggregated dynamics of strategy distributions over time. In each plot, the x-axis represents the periods over time; the y-axis represents the percentage of agents using a specific strategy.

In the *Control* condition, the system initializes at the even distribution of the five strategies. Since agents have no preferences and no history for them to make calculated decisions, they play randomly at the very beginning. Later, some strategies gain popularity among agents while others lose. Eventually; each strategy's share becomes stable after hundreds of

Table 1: The list of parameters and their default values.

Parameter		Description	Default Value
Social Dilemma Payoffs	*T*	An agent's payoff when playing *D* against *C*.	1.0
	R	An agent's payoff when playing *C* against. *C*.	0.6
	P	An agent's payoff when playing *D* against *D*.	0.2
	S	An agent's payoff when playing *C* against *D*.	0.0
Commitment Setups	*e*	Cost of making a commitment.	0.1
	w	Penalty to defect after a commitment.	0.5
Individual Differences in Propensity to Cooperate		Strongly prefer to cooperate (20%).	[0.6, 0.9, -0.6, -0.9, 0]*
		Moderately prefer to cooperate (20%).	[0.2, 0.5, -0.2, -0.5, 0]
		Neutral (20%).	[0, 0, 0, 0, 0]
		Moderately prefer to defect (20%).	[-0.2, -0.5, 0.2, 0.5, 0]
		Strongly prefer to cooperate (20%).	[-0.6, -0.9, 0.6, 0.9, 0]
Bounded Rationality	*m*	The number of past interaction agents can observe.	5
	θ	Probability of making random decisions.	0.05
Simulation Settings	*N*	The population size of the agents in a simulation.	50
	I	The number of the interactions simulated in a single simulation run.	1,000

Note. * the idiosyncratic payoffs corresponding to strategies [*COM_C, C, D, COM_D, COM_O*], similarly hereinafter.



a. The dynamics of the proportions of agents using each strategy without considering individual preferences in cooperation (**Control**).

b. The dynamics of the proportions of agents using each strategy with considering individual preferences in cooperation (**Treatment**).

Figure 4: The dynamics of strategies under two experiment conditions.

interactions. *COM_C* becomes the strategy adopted by most individuals in the long run. Roughly 28% of agents would use it. *COM_O* is the second most popular one, which is used by 25% of agents. The third popular strategy is *COM_D* (21%). The fourth is *D*, and the fifth is *C*. They account for about 13% and 12% of agents, respectively.

The initial phase under the *Treatment* condition is different (Fig. 4.b) since people tend to play their favorite strategies at the beginning when considering their individual preferences. The long-term strategy choices are different. *COM_D* is the most popular condition, which is played by about 35% of agents. *C* holds the second place with about 29% of agents using it. The other three strategies are adopted by a similar amount of agents, which are 13% for *COM_C*, 12% for *COM_O*, and 11% for *D*.

Table 2: The proportions of agents using each strategies.

	Control	*Treatment*
COM_C	≈ 28%	≈ 13%
COM_C + *C* (Lean to defect)	≈ 40%	≈ 42%
COM_D	≈ 22%	≈ 35%
COM_D + *D* (Lean to defect)	≈ 35%	≈ 46%

Regardless of the differences in the early simulation initialization phase, there are several critical differences in the long-term dynamics (Tab. 2). It seems that the commitment's effects in promoting and maintaining cooperation are undermined. The most uncooperative and antisocial strategy (*COM_D*: making commitments but never fulfilling them) becomes the top choice. While about the same amount of people choose to be cooperative by playing *COM_C* and *C*, much more people lean to be uncooperative unconditionally (*COM_D* + *D*: 46% vs. 35%). The conditional cooperative individuals (using *COM_O*) are out of the game (25% vs. 12%). However, commitment still has some positive effects by preventing those who prefer to cooperate from adopting uncooperative strategies. Note that the number of agents who use *C* is much higher under the treatment condition. It helps maintain the bottom-line cooperation in a community.

Therefore, we can answer the $\mathbf{RQ_1}$ as follows:

> *The commitment's positive effects on promoting cooperation are undermined but still exist. The amount of individuals who take cooperative strategies remains at a similar level (42% vs. 40%), but more individuals may choose the antisocial strategy and lean to defect unconditionally (35% vs. 22%). In general, commitment's effects are exhibited in the form of preventing people who prefer to cooperate from taking uncooperative strategies.*

Experiment 2 Results & Findings

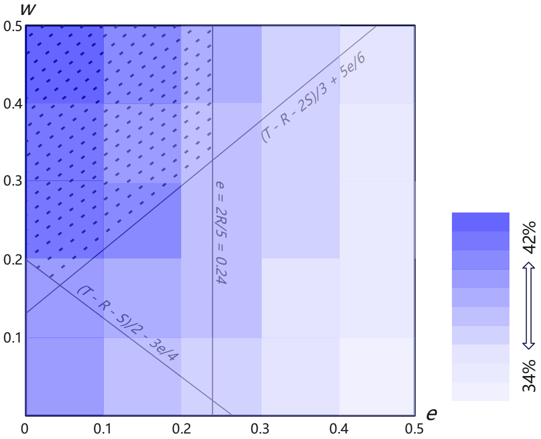


Figure 5: The average proportions of using strategies leaning to cooperation under different setups of commitment.

We use a heatmap (Fig. 5) to visualize the average proportions of individual agents using strategies *C* or *COM_D* under the 25 combinations of $< e, w >$. The results are straightforward. The top left corner ($e = 0.1$ and $w = 0.5$), which indicates the combination of smallest cost of commitment and highest punishment, records highest proportion of agents using *C* or *COM_D* (42%). The proportions gradually decrease in both directions until the bottom right corner, where 34% of agents use *C* or *COM_D*. Only one exception happens at $e = 0.2$ and $w = 0.3$. From Fig. 5, it is easy to conclude that the effects of different setups of commitment are in a similar direction with prior literature such as Han et al. (Han et al., 2012). However, recall that the simulations initialize from a situation where about 45% of agents play *C* when considering individuals' diverse preferences. Even the best case with the commitment $< 0.1, 0.5 >$ has 3% loss (42% vs. 45%) regarding the proportion of agents who play cooperative strategies. In this sense, the effect of the commitment mechanism is **not guaranteeing** a majority of members to be cooperative, but **preventing** members from changing to uncooperative ones, which is also consistent with $\mathbf{RQ_1}$'s findings.

Therefore, we can answer the $\mathbf{RQ_2}$ as follows:

> *Different setups of commitment still have impacts on members' long-term strategic choices with diverse individual preferences. In general, the smaller cost for making commitments and the larger penalty for failing to fulfill commitments would lead more individuals to be cooperative but cannot guarantee a majority to use strategies leaning toward cooperation. Compared with the literature without considering individual preferences, we obtain similar yet weaker results.*

Discussion

This article reports on our agent-based modeling and simulation efforts for investigating the complicated interrelations between commitment and cooperation with special consideration of diverse individual propensity to cooperate. We combine multiple theoretical insights and empirical evidence to design the ABM and run extensive simulation experiments with it. Our results reveal that: when considering individual preferences, (1) commitment's positive effects on promoting cooperation are undermined but still exist and exhibit in the form of preventing people of goodwill from taking uncooperative strategies; (2) different setups of commitment mechanisms still matter, and smaller cost of making commitments and larger penalty for failing to fulfill commitments are desirable. Our results inform decision-makers to properly evaluate their commitment mechanisms' effectiveness with the characteristics of their communities' members. The recent flourish of automated preference inference techniques, e.g., Houlsby et al. (Houlsby, Hernández-Lobato, Huszár, & Ghahramani, 2012), significantly reduces the cost of performing such evaluations. Decision-makers could reuse our ABM model and combine it with the individual preferences inferred from members' digital traces to run mechanism evaluations before launching in their communities. Future work will continue to investigate the complex dynamics resulting from the interaction among commitment, cooperation, and other social and psychological factors under the highly-extensible ABM we developed and validate the analytical results with empirical and lab studies.

Acknowledgements

This research was partially supported by the National Natural Science Foundation of China under grants 62172049 and 62076232.

References

Ajmeri, N., Guo, H., Murukannaiah, P. K., & Singh, M. P. (2020). Elessar: Ethics in norm-aware agents. In *Aamas* (Vol. 20, pp. 16–24).

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396.

Back, I., & Flache, A. (2006). The viability of cooperation based on interpersonal commitment. *Journal of Artificial Societies and Social Simulation*, *9*(1), 12. Retrieved from https://www.jasss.org/9/1/12.html

Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: a meta-analytic review of social dilemmas. *Psychological bulletin*, *137*(6), 881-909.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl_3), 7280-7287. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.082080899 doi: 10.1073/pnas.082080899

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology*, *65*(1), 17–28.

Chen, X.-P., & Komorita, S. S. (1994). The effects of communication and commitment in a public goods social dilemma. *Organizational Behavior and Human Decision Processes*, *60*(3), 367–386.

Corbett, E., & Le Dantec, C. A. (2018). Going the distance: Trust work for citizen participation. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3173574.3173886 doi: 10.1145/3173574.3173886

Fudenberg, D., & Levine, D. K. (1998). *The theory of learning in games* (Vol. 2). MIT press.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. MIT press.

Gibbons, R., & Gibbons, R. (1992). A primer in game theory.

Han, T. A. (2016). Emergence of social punishment and cooperation through prior commitments. In *Proceedings of the thirtieth aaai conference on artificial intelligence* (pp. 2494–2500).

Han, T. A., Pereira, L. M., & Santos, F. C. (2012). The emergence of commitments and cooperation. In *Proceedings of the 11th international conference on autonomous agents and multiagent systems - volume 1* (p. 559–566). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Hauser, O. P., Hilbe, C., Chatterjee, K., & Nowak, M. A. (2019). Social dilemmas among unequals. *Nature*, *572*(7770), 524–527.

Houlsby, N., Hernández-Lobato, J. M., Huszár, F., & Ghahramani, Z. (2012). Collaborative gaussian processes for preference learning. In *Proceedings of the 25th international conference on neural information processing systems - volume 2* (p. 2096–2104). Red Hook, NY, USA: Curran Associates Inc.

Jones, G. (2008). Are smarter groups more cooperative? evidence from prisoner's dilemma experiments, 1959–2003. *Journal of Economic Behavior & Organization*, *68*(3-4), 489–497.

Kazil, J., Masad, D., & Crooks, A. (2020). Utilizing python for agent-based modeling: The mesa framework. In R. Thomson, H. Bisgin, C. Dancy, A. Hyder, & M. Hussain (Eds.), *Social, cultural, and behavioral modeling* (pp. 308–317). Cham: Springer International Publishing.

Kortenkamp, K. V., & Moore, C. F. (2006). Time, uncertainty, and individual differences in decisions to cooperate in resource dilemmas. *Personality and Social Psychology Bulletin*, *32*(5), 603–615.

Lave, L. B. (1962). An empirical approach to the prisoners' dilemma game. *The Quarterly Journal of Economics*, *76*(3), 424–436.

Marwell, G., & Oliver, P. (1993). *The critical mass in collective action*. Cambridge University Press.

Moisan, F., ten Brincke, R., Murphy, R. O., & Gonzalez, C. (2018). Not all prisoner's dilemma games are equal: Incentives, social preferences, and cooperation. *Decision*, *5*(4), 306.

Morrison-Smith, S., & Ruiz, J. (2020). Challenges and barriers in virtual teams: a literature review. *SN Applied Sciences*, *2*, 1–33.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560-1563. Retrieved from https://www.science.org/doi/abs/10.1126/science.1133755 doi: 10.1126/science.1133755

Olson Jr, M. (1971). *The logic of collective action: Public goods and the theory of groups, with a new preface and appendix* (Vol. 124). Harvard University Press.

Ostrom, E. (2003). Toward a behavioral theory linking trust, reciprocity, and reputation. *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, *6*, 19–79.

Pearce, J. L., Branyiczki, I., & Bigley, G. A. (2000). Insufficient bureaucracy: Trust and commitment in particularistic organizations. *Organization Science*, *11*(2), 148–162.

Ren, Y., & Kraut, R. E. (2014). Agent-based modeling to inform online community design: Impact of topical breadth, message volume, and discussion moderation on member commitment and contribution. *Human–Computer Interaction*, *29*(4), 351-389. doi: 10.1080/07370024.2013.828565

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Sasaki, T., Okada, I., Uchida, S., & Chen, X. (2015). Commitment to cooperation and peer punishment: Its evolution. *Games*, *6*(4), 574–587. Retrieved from https://www.mdpi.com/2073-4336/6/4/574 doi: 10.3390/g6040574

Simon, A. F., Fagley, N. S., & Halleran, J. G. (2004). Decision framing: Moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making*, *17*(2), 77–93.

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90.

Wang, Y., & Redmiles, D. (2016a). Cheap talk, cooperation, and trust in global software engineering: An evolutionary game theory model with empirical support. *Empirical Software Engineering*, *21*, 2233–2267.

Wang, Y., & Redmiles, D. (2016b). The diffusion of trust and cooperation in teams with individuals' variations on baseline trust. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing* (p. 303–318). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2818048.2820064` doi: 10.1145/2818048.2820064

Young, H. P. (1998). *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press.

Zeigler, B. P., Praehofer, H., & Kim, T. G. (2000). *Theory of modeling and simulation*. Academic press.

Zheng, J., Veinott, E., Bos, N., Olson, J. S., & Olson, G. M. (2002). Trust without touch: Jumpstarting long-distance trust with initial social activities. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 141–146). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/503376.503402` doi: 10.1145/503376.503402