

# A Computational Model of the Situationist Critique

Yang Renjie. Capital Normal University

DOI: 10.47297/wspciWSP2515-469901.20200401

## 1. Introduction

Virtue ethics is commonly regarded as a third approach to normative ethics. Different from deontology and consequentialism, virtue ethics is established on the idea that the most foundational concepts of morality are virtues and vices. A virtue is a personal character trait that depicts one's stable disposition to act, reason, or feel in certain way.<sup>[1]</sup> According to virtue ethicists, all the other prevailing moral concepts, such as moral duties or good consequences of actions, are only secondary in the sense that they should be explained in terms of virtues and vices. In this way, virtue ethics provides an alternative moral theory about what is good or bad, what is right or wrong. Since the publication of Anscombe's famous paper "Modern Moral Ethics" in 1958,<sup>[2]</sup> virtue ethics has attracted great attention from moral philosophers. It seems to have the potential to overcome some of the major issues in the other two normative moral theories, such as the dependence on the concept of a "divine lawmaker" in making moral judgements as complained by Anscombe. By shifting our attention from abstract entities such as moral duties or rules of actions to concrete but fundamental questions about what kind of life we should live, virtue ethics revives the ancient tradition of morality in both the Western and Eastern worlds. However, there are many different versions of the

theory and the controversy about what kind of theory we should establish is far from settled. What is the nature of a virtue and how should we understand it? How many types of virtues are there, and how are they related to practical benefits? Can virtues be learned or cultivated? Can we establish a satisfactory theory of morality based on virtues and vices? These are important problems that virtue theorists need to resolve.

In recent years, a group of philosophical situationists raised a serious line of objection against the basic assumptions of virtue ethics. Philosophers such as Gilbert Harman,<sup>[3]</sup> John Doris,<sup>[4]</sup> and Maria Merritt<sup>[5]</sup> draw attention on the empirical studies on moral behaviors and personality traits from the social psychology literature. One of the most cited experiment is Stanley Milgram's famous obedient experiments.<sup>[6]</sup> In this series of experiments, the psychologists told the subjects that they were conducting an empirical study about learning and punishment, but their real purpose was to test the subjects' obedient behaviors under the pressure of authoritative figures. The subjects were asked to electrically shock the confederate "learners" when they made a mistake in the learning and testing tasks. The experiment results show that over 65% of the subjects did not insist on fighting against this apparently immoral command, which was much higher than the percentage expected by both experts and laypeople back in the 1960s. In another experiment, Alice Isen and Paula Levin.<sup>[7]</sup> tested the subjects' helping behavior by arranging a confederate to drop a pile of files near them. Before that happens, the psychologists asked the subjects to find a coin in a public pay phone, and manipulated their mood by changing the difficulty of finding the coin for different subjects. It turned out that for those who did not find the coin and thus had a more negative mood, the rate of helping behavior dropped from 87.5 percent to 4 percent.

Philosophical situationists believe that these and many other empirical studies shatter the theoretical foundation of virtue ethics. Milgram's experiment and related studies seem to show that human moral behaviors are not caused by character or personality traits, but rather by the environmental and situational factors. Virtue ethicists try to build their ethical theory on the concept of virtue as a stable character trait, but the empirical evidence does not support the stability of the virtues under different environmental circumstances. Without an ideal of a stable virtue that

takes effect in most of the moral behaviors, the success of a normative theory of virtue ethics is hard to imagine. In addition, Alice Isen's experiments and related results show that many situational factors usually affect moral behaviors in an unconscious way. Following the Aristotelean eudaimonic tradition, one of the central beliefs of virtue ethics is that morality can be cultivated through moral education. It is believed that moral education is a process of rational and conscious training of moral characters rather than the inculcation of a set of behavioral rules. This belief relies on the causal efficacy of the virtues. If the human moral behaviors are determined mostly by unconscious processes, then the conscious training of the virtues would be in vein, because no matter how good a person gets to know the virtue consciously, the moral character will not take effect in their behaviors. Even though the situationists did not refute virtue ethics decisively, the burden of proof shifts to the virtue ethicists.

The virtue ethicists respond differently to the challenge raised by the situationist critique. One common response is that the empirical results cited by the situationists have nothing to do with the normative theory of virtue ethics,<sup>[8]</sup> because the Aristotelian virtue ethicists have already made the claim that the true virtues that are stable are very rare in reality. Some other virtue ethicists admit that the experiments in social psychology are relevant, but disagree with the situationists about whether they show that globally stable virtues don't exist.<sup>[9]</sup> Still other virtue ethicists bite the bullet and admit the situationist challenge, but continue to develop a modified version of virtue ethics theory that could avoid the difficulty.<sup>[10]</sup> In this paper, I will present another way to respond to the situationist critique that is different from all of the above ones. The basic idea is to draw attention on some of the phenomenon one might encounter in typical moral decision-making scenarios that are often omitted or ignored. The discovery of these phenomenon takes advantage of the power of computation models to illustrate emergence phenomenon in complex systems. Once this additional information is taken into account, we will be able to see more possible ways for virtue ethics to meet the situationist challenge. The structure of the paper is organized as follows. After the introduction to an agent-based model for the virtue of justice in Section 2, I will establish a novel argument that defends the virtue ethics against the situationist critique in section 3,

and describe two simulation experiments to support the argument. Section 4 continues to reflex on the assumptions made in the computation model as well as the implications of the simulation results on virtue ethics. I will also reply to some possible objections. The last section concludes.

## 2. An Agent-Based Model of Justice

In the last several decades, the rapid development of artificial intelligence technologies had greatly expanded the toolbox for social sciences and humanity researches. The agent-based modeling method is one of them. While it is developed and widely used in many social science areas such as economics and sociology, the agent-based modeling of ethical theories is still rare. One notable exception is [11], but the focus is on consequentialism exclusively. In this section, I will propose an agent-based model of the virtue of justice.

Agent-based modeling is especially good at modeling and representing complex systems. A typical complex system consists of a large amount of relatively small components, each one of the components is subject to the influence of both the other components and the larger structures that are made of them. Many aspects of human society satisfy this defining characteristic of a complex system. From this perspective, social phenomena can be viewed as the result of the causal interactions of a large number of individual creatures who can make decisions and conduct actions independently. For example, suppose every participant in a free market largely follows the requirements of the rational decision-making rules, then at the global level the free market will exhibit equilibrium states in which market supply and demand balance each other. This macro-level phenomenon emerges from the micro-level interactions of the individual components. Agent-based modeling is a modeling technique in the area of complex science. It is especially useful in modeling complex systems as described above. The idea is to model in a virtual agent the exact decision procedures and mechanisms of each individual in the micro-level description of the system, and use computer simulations to illustrate how they interact with each other and produce higher level phenomena. One advantage

of the agent-based modeling methodology is that the concrete simulations of the dynamics of a complex system have the power to uncover possible emergent phenomenon that goes beyond the reach of direct human intuitions. As a notable example, the famous Shelling's study<sup>[12]</sup> of racial segregation shows that the global level of racial grouping can happen in cases where individuals are nowhere near racists. As we shall see, the argument I propose in this paper takes advantage of this property of agent-based modeling methodology.

The modeling of virtue ethics demands a setup of a concrete social situation. To avoid reinventing the wheel, I take the well-known El Farol model as the basis model, and adapt it to model the virtue of justice. Similar ideas could be applied to other types of virtues, but this example is enough for a demonstration of the concept. In his original description of the El Farol model, Arthur<sup>[13]</sup> used a beer bar named El Farol in Santa Fe, New Mexico as a prototype. The bar is famous for its Ireland music performance every Thursday. On that day, there are 100 local customers who usually stop by the bar and enjoy the music. The bar is too small to host all the 100 customers. It can only accommodate 60 people at one night. If the number of people went beyond 60, the bar would become too crowded and everyone's night would be ruined. To solve this dilemma, the local newspaper publishes the number of customers in the bar the week before. It is up to each person to determine for themselves if they should go to the bar according to the data of the past several weeks. Arthur designed a model for this scenario. Figure 1 shows its implementation in the Netlogo software, where the upper right section represents the El Farol bar.

As one of Plato's four cardinal virtues, justice is related to the fairness in the distribution of a public resource. The El Farol model depicts a problem about the fair distribution of the 60 seats in the bar among the 100 customers. The overall distribution is a collective

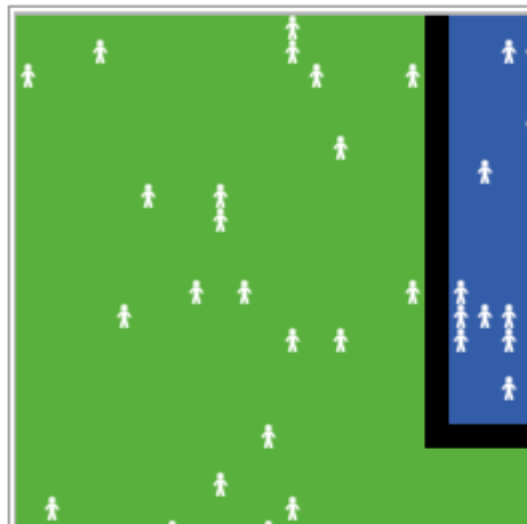


Fig.1 The El Farol Model

result of the decisions made by each individual customer. Since they all love the music and the beer, the natural choice for each customer is to go to the bar this Thursday. But the bar would be overcrowded if they all followed their natural choice. The result is a loss for everyone. On the other hand, if no one came to the bar because of the above line of reasoning, then the public resource in the bar would be wasted. The ideal solution is precisely 60 customers attending the bar. As part of the model setting, no collaboration of any kind is allowed between any group of customers. They must make their own decisions independently to make sure they reach exactly the golden number of 60. The virtue of justice is a character trait of the individual customers. It is expected that by possessing the virtue of justice individually, the group of 100 customers will arrive at the ideal macro-level solution in a stable manner.

I propose the following definition of the virtue of justice in the El Farol model describe above: we say that an agent has the virtue of justice if she does not go the bar when she expects that the bar will be overcrowded. To see how this definition captures the philosophical conception of the virtue of justice, I will elaborate on several implications of the definition on virtue ethics. First, since everyone loves the music in the bar, and everyone knows that no one will enjoy the music if too many people come to the bar simultaneously, we tend to say that those who pass their opportunity at appropriate time demands only a fair share of the public resource, thus possess the virtue of justice. Second, since both consequentialism and deontology can have an account of moral virtues, the defining feature of virtue ethics is to take the concept of virtue as the theoretical foundation. Therefore, one demand of virtue ethics is that the correct definition of a virtue should not be the maximization of the benefit or utility of the moral behavior. Consider the situation in which going to the crowded bar and staying at home have the same utility for any individual. In this case, it follows from the principle of maximal utility that the best decision is to go to the bar every week no matter what happens, because the utility of going to the bar is no less than staying at home. This decision apparently violates the demand of justice. Therefore, at the individual level the virtue of justice is different from the maximal benefit of the behavioral consequence. Third, the virtue of justice cannot be reduced to the maximal utility at the global level either.

Consider the situation where there are more than 60 person who always come to the bar, perhaps because of a consequentialist mind. Then the global level behavioral result is always crowded, regardless of what the other people do. In this case, for any given agent, having a virtue of justice will not change anything, but it still has a moral value in itself. This is consistent with the virtue ethicists' intuition about the concept of virtue. Fourth, the above definition of justice has a corresponding vice as an opposite character trait. If there is someone who insist on going to the bar even when she has perfect good reasons to believe that the bar will be crowded, we tend to say that this person not only lacks the virtue of justice, but actually possess a vice of intentionally sabotaging everyone's plan. Fifth, our definition of justice takes into account the intellectual virtue *phronesis*, or practical wisdom. Processing the virtue of justice does not guarantee a morally good behavior, because the agent might underestimate the number of people who are willing to go to the bar. Consider a virtuous person who is always motivated to be just, but not smart enough to predict correctly the number of people in the bar even once. Then this person does not act justly at all from other people's perspective, because she would always go to the bar regardless. This example illustrates the importance of practical wisdom for the virtues to take effect in moral actions, a point which is usually emphasized by virtue ethicists. Lastly, the above example of virtue without wisdom also exhibits the relevance of inner motivations. Even though the unwise person in the example never conduct any just behavior, there is no doubt that she has something admirable about her character trait, because she has the spirit to not give up in making effort to act justly. The necessity of virtuous motivation for a virtuous person is also emphasized by many virtue ethicists. In short, there are enough reasons for us to believe that the proposed definition of the virtue of justice is an appropriate manifestation of the virtue ethicists' conception of virtue in the specific social situation of the El Farol problem.

The other component in the El Farol model is the decision mechanism of the agents. In Arthur's original model, the agents have the capacity to learn from the past experience, namely the last several weeks' record of the number of customers in the bar. Every agent is randomly assigned a set of predictors, which could be used to predict the number of customers in the bar in the current week. The imple-

mentation of the predictor is a simple linear model with the number of customers in the past weeks as independent variables. Therefore, each predictor is represented as a set of parameters of the linear model. The agents have a limited memory of the past experience with a fixed length, which keeps updating as time goes. In every week, each agent chooses the best predictor in its set of strategy, which is defined as the predictor that has the highest correct rate when tested on the agent's memory of the past experience. The optimal predictor might change when the memory gets updated. The agents then make a prediction on whether or not the bar will be overcrowded, and decide on which action to choose: going to the bar or staying at home. I keep this implementation of bounded rationality in the justice model, but only take it to be the rational component of the cognitive model for the agents. There are two additional components in the model, one for the unconscious cognitive processes of the agents, another for the situational or environmental factors. They will be described in detail in the next section.

### **3. Experimental Design and Results**

The challenge raised by the situationist critique can be summarized as follows. Situationists propose that most of the virtuous behavior we witness in our daily life are caused by some seemingly unimportant non-ethical environmental or situational factors. Also, most of these situational factors take effect unconsciously. They cite many empirical studies of social psychologists as their evidence. Situationists derive from this observation to the conclusion that a) in reality there is unlikely such thing as stable character traits under distinct environments or situations, therefore the concept of virtue on which virtue ethics establish their brand of normative ethics is shattered; b) the effort to cultivate such stable character traits through conscious moral cultivation and education is most likely to fail, because the situational factor will sabotage the conscious demand of virtue in moral behaviors.

My response to the above two objections is this. Situationists made the right observation in social psychology studies, so their premises might be true. But the



two conclusions they draw about the empirical results' implications on virtue ethics do not hold. In other words, they are wrong about what the empirical results mean for virtue ethicists. My argument goes like this:

(1) We can model the situational factor and the character factor computationally, and experiment on how each factor would change the virtue social setting.

(2) The computational model shows that even a small portion of character factor would significantly change the aftermath of the moral behaviors.

(3) The computational model also shows that even when most of the decisions on moral behaviors are made unconsciously for all the people, conscious virtue learnings can still have a significant effect on the behavioral results in the social setting.

(4) Premise 2 indicates that we simply cannot directly remove the concept of virtue out of our moral thinking, as the situationists suggest. Even though most of the character traits in reality are not stable, it doesn't pose a threat for virtue ethics because virtues can take effect even when they are highly unstable. Our moral theory should take this effect into account, and that requires the concept of virtue.

(5) Premise 3 shows that the effort of moral education and virtue cultivation is not in vein. Even if the causal power of the character trait is mostly blocked by unconscious factors, it can still significantly affect the behavioral results. That means the aim of moral improvement toward the ideal, rather than directly achieving the ideal, is not hopeless, but quite practical, worthwhile, and achievable.

(6) Therefore, the situationist critique of virtue ethics does not hold.

In the rest of this section, I will describe two computational experiments that support premises 2 and 3. Let's start with premise 2. As pointed out by the situationists, human moral behaviors are mostly influenced by the environmental factors. I add two more coefficients to model their effects: an environmental difficulty coefficient  $\alpha$  representing how hard the environment goes against the agent's will, and a deterministic coefficient  $\beta$  representing the agent's will power to overcome the adversarial environment and to follow the demand of the virtues. I assume the environment always has a tendency to force every agent to go to the bar, contrary to the intention of the just agents who predict that the bar will be crowded.  $\alpha$  refers to the probability that the environmental factor makes the agents go to the bar. If  $\alpha$

is 1, then the environment always requires the agents to go to the bar; if  $\alpha$  is 0, then the environment always requires the agents to stay at home. Every time an agent makes a decision, it follows the instruction of the rational component of its cognitive model with probability  $\beta$ , and subject to the demand of the environmental factor with probability  $1 - \beta$ . For each possible result of the behavior, assign a set of rewards as follows: when the bar has less than 60 agents, those who goes to the bar get reward 1, while those who does not get reward 0.1; when the bar accommodates more than 60 agents, everyone gets a reward of 0.1. This reward system helps separate the virtuous behaviors from the consequentialist behaviors. Since going to the bar is always at least as good as staying at home no matter how crowded the bar is, a consequentialist will always go to the bar based on the dominance decision rule. In contrast, a virtuous agent will refrain from going when she foresees a bar full of people.

The following computational experiment is conducted under the model setting describe above. Each agent keeps a memory of the past 5 weeks' attendances, and holds a strategy set with 10 candidate predictors. Let the rational decisions of the agents follow the virtuous decision rule. In other words, when an agent consults its rational strategy and predicts that the bar will be crowded, then the agent will choose not to go. Someone who make the opposite decision cannot be regarded as processing the virtue of justice. For a given value of  $\alpha$  and  $\beta$ , let the agents make a choice between the rational decision and the environmental dictation with probability  $\beta$  and  $1 - \beta$  respectively for 1000 times. The probability of going to the bar under the environmental control is set to  $\alpha$ . Repeat the simulation 10 times, and record the average value of the crowded rate of the bar. Figure 2 shows the simulation results. We can observe two facts from this experiment. First, when the behaviors of the agents are determined by their rational predictions and virtuous decision rules, the number of agents in the bar fluctuate around a number close to 60. This well-known emergent phenomenon yields additional implications on virtue ethics in the justice model. As part of the model setting, the behavior of every individual agent is caused by two factors: the rational or virtuous cognitive component and the environmental or situational factor. The causal effect of both of the two factors can be explicitly measured by the overcrowding rate of the bar. When the agents

are controlled mainly by the rational strategies, the overcrowding rate is lower. When they bend under the pressure of hard environment, the overcrowding rate is higher. Second, even when the environment is very hostile while the will power of every agent is very weak, there is still a significant improvement on the overcrowding rate. This fact is usually overlooked by the philosophical situationists, perhaps because its discovery requires a level of detail that is best illustrated by a computational simulation. It shows that one cannot ignore the causal effect of virtuous character traits even when they are subject to severe interferences by the environmental factors. The situationists might be correct when they observe that our moral behaviors in reality are mostly caused by environmental factors. But that does not mean we need to abandon the conceptions of virtues and vices, because they are necessary for us to account for the significant overall effects exhibited in fact two which is caused by a small contribution of the virtuous factor.

To support premise 3, we need to model both the unconscious mechanism of the decision procedure for each agent and the conscious virtue learning mechanism among them. The unconscious component of the justice model follows Erev and Roth's idea of reinforcement learning.<sup>[14]</sup> The agents in the model can either go to the bar or stay at home. Set an impulse value for each of the two options and update them as the agents make decisions by accumulating separately the rewards they collect. The reward setting is the same as the one described before: 1 for going to the bar when it turns out to be spacious and 0.1 for all the other combinations. The possibilities for an agent to choose between the two options is set to be proportional to their corresponding impulse values. Situationists challenge the idea of virtue cultivation and moral education. One mechanism of virtue learning commonly suggested by the virtue ethicists is learning from admirable role models. The purpose is not to mimic exactly what the role model does in every circumstance, but rather to deliberate about the reasoning required for the role models to act virtuously. The learning process in the justice model follows a similar idea. An agent can only learn another agent's prediction strategy in its conscious component, but not their exact actions or anything in their unconscious components. A learning step is defined as the procedure of deleting one of the learner's own predictor in its strategy set and add the one in the target's strategy set that predicts best in the current step.

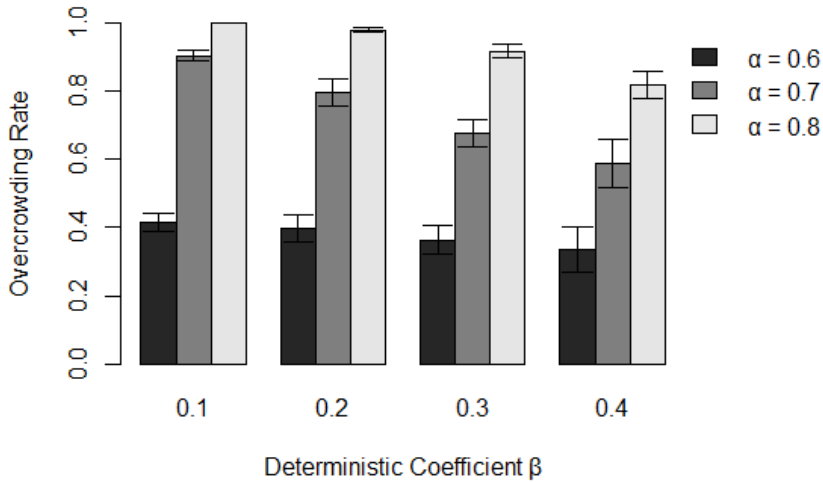


Fig. 2 Experiment 1 Simulation Results

With the unconscious component and the learning procedure added to the model setting, the next experiment tests the effect of virtue learning when all the agents are subject to the influence of both their conscious and unconscious cognitive processes. Let  $\gamma$  denotes the probability that the agent follows the decision made by the rational component rather than the unconscious one. For every time point, let one of the agents who violates the virtue of justice most frequently learn from one of those who violates the least as described above, repeat this learning procedure for 10 times. The simulation result is shown in Figure 3. Remarkable, even when most of the conscious decisions of the agents are interfered by the unconscious impulses, the learning procedure still improves the overcrowding rate and the mean reward of the agents. Note that the learning process is designed to be relatively weak: only 10 percent of person learn from 10 percent of the strategies adopted by the virtuous agents. Contrary to the expectation of the situationists, the effort to cultivate virtuous traits through conscious moral learning does not fail, even though the trait in question is unstable all along.

The above result indicates a distinction between ideal virtue and virtue improvement. The latter does not have to be ideal, no matter how small it is. As long as it is not nil, it is morally significant. In order to show this distinction, the computational experiment is intentionally designed to magnify its practical significant. The ideal virtue is mostly favored by the Aristotelians. If we use Aristotelean vir-

tue ethics only as an ideal, and admit virtue improvements as part of virtue, then the situationist is mistaken about the implication of their second observation. Virtue ethics does not require the actual achievement of the ideal virtue. It can serve as a conceptual ideal that is necessary for capturing the concept and direction of practical efforts. Since the situational factors will not completely diminish all the causal power of the character factors, the remaining causal effect of virtue improvement can still be significant, which keeps the practicality of the moral ideal. The ideal virtue is morally significant because it is the aim of virtue improvement. The concept of virtue improvement does not subject to the critique posed by situationists, because it only requires a non-zero possibility of happening to maintain its moral significance and practically significant. Therefore, both the ideal virtue and virtue improvement are legitimate concepts for moral theorists.

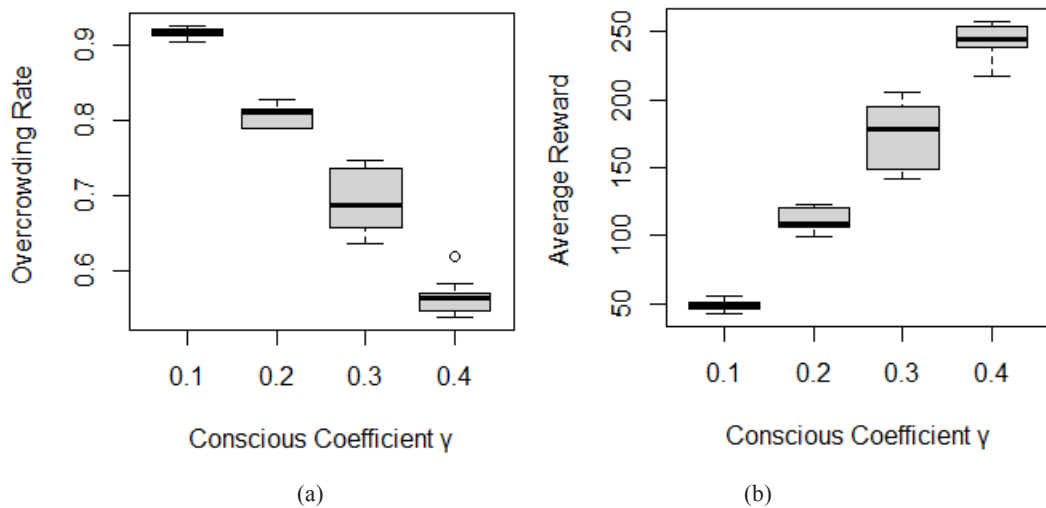


Fig. 3 Experiment 2 Simulation Results

#### 4. Limitations and Objections

A common attack against any agent-based modeling study or any computer simulation research in general is to point out the unrealistic nature of the model. To turn on the simulation engine and keep it running, the researchers must simplify the issue at hand extensively and make up some pretend values for the model pa-

rameters. These moves will no doubt negatively affect the validity of the model. Computational social scientists usually counter by saying that all models are wrong, but some are useful.<sup>[15]</sup> I don't find this response particularly satisfactory. The validity of a model depends on whether the model and the target system have some relevant structures that are similar in a meaningful way. I will discuss four aspects of the model assumptions that are crucial for the justice model: the reward setting, the definition of justice, the conscious and unconscious cognitive components, and the learning procedure.

In the computational experiments, the rewards are used to quantify the action utility of every agent. They should be distinct for different agents in reality. The reward setting in the justification model is an uncommon one: 1 for uncrowded and go, and 0.1 for all the other possibilities. Why are staying at home and being in a crowded bar equally unpleasant? After all, people would prefer staying at home if the bar is crowded, which is exactly how the original El Feral model is set up. But in that case, there would be no behavioral distinction between virtuous people and profit-driven ones. In fact, if an agent's only purpose is to maximize its reward, then she should choose to go home whenever she predicts that the bar will be full, because it is a more profitable option. In that case, any virtuous behavior manifested in the model could be explained by the agent's consequentialist motive. The reward setting of the justice model is designed to avoid this situation, so that we can expose the effects of the virtue.

Despite the lengthy justification of the definition of justice in section 2, one might still find it lacking at some point. If we think of the seats in the bar as a kind of public resource, then wasting it intentionally might be a vice. Why isn't going to the bar when there are still seats available part of the virtue of justice? The problem is similar to the one I just mentioned, namely if it is then we can't separate virtuous people from profit-driven ones because they will all behave the same way. This response raises another worry: doesn't the definition of justice adopted in the computation model violate our intuition, at least in this aspect? The answer might be yes, but it won't affect the general argument. In fact, even if the pattern of behavior one takes as virtuous is completely artificial, the learning mechanism in the

justice model still works as long as the agents aim to achieve that behavioral pattern. We can use our intuitive definition of justice instead. I am narrowing it only to make its behavioral effects more explicit.

The cognitive components might be a weak spot of the computational model of justice. Both the conscious and unconscious models are wildly simplified and unrealistic. The main reason I am still using them is that they are well known in the literature. I admit that they could be improved by updating to an implementation of a more comprehensive cognitive architecture of human decision-making processes with empirical validity. For example, we may replace Arthur's original linear models of predictors by a more sophisticated deep neural network predictor. Another useful improvement is to add an affective component that is specialized for modeling emotions and feelings. In Milgram's obedient experiments, if we consider the internal struggle experienced by the subjects when they follow the order to shock the confederates, we might have a different opinion about the overall effects of the situational factors. The agents' internal feelings and other affective processes are not reflected in the current model of justice. We need further studies to know if these improvements would change the experimental results.

There is one subtle objection against the learning procedure in the model. Imagine the learning ability of the agents are so great that ultimately, they all think exactly like the most virtuous agent. When that happens, all the agents will make the same decision, which means they all either go or not go simultaneously. This is the worst possible distribution of the public resource. Ironically, all of them are as virtuous as possible. One response to this objection is to think of an opposite situation where the agents learn from the most vicious one instead. They are all equally evil, and they get what they deserve. The key to this problem is that the El Farol setup requires a reasonable level of heterogeneity among the agents. Since the unconscious component of the cognitive model and the probabilistic nature of the decision procedure provide enough differences among the agents, this objection is avoided in the model of justice.

One final objection I would like to consider concerns the foundational role of the virtues. One of the defining features of virtue ethics as a normative theory is that it takes the concept of virtue as basic. All the other moral concepts such as

good or bad actions are defined in terms of it. However, in the argument proposed in this paper, the moral significance of the virtues is justified by noting their macro-level behavioral consequence, namely the overcrowding rate of the bar. Does that support the consequentialist idea by defining virtues in terms of favorable consequences? Consider the cases where the environment is so hostile that virtue factors have no influence on the consequences at all. For example, suppose everyone is forced to go to the bar no matter how virtuous they are. Virtue ethicists can still argue that the virtues are valuable in themselves. It is just that they didn't manifest in the consequence. The difference in consequences is used as an illustration of moral significance of virtues, not as a definition.

What do all these mean for virtue ethics? First, considering the various limitations of the computation model of justice, the argument proposed in this paper does not settle the debate between virtue ethicists and the situationists. Second, even though the argument is not conclusive, it paves the way for a promising response to the situationist critique that should be taken seriously. The emergence phenomenon uncovered by the simulation experiments show how subtle the causal effects of character traits can be. They also help clarify the relevance of the distinction between the concept of ideal virtues and virtue improvements, a distinction that is often ignored in the situationism debate. Third, aside from empirical studies in psychology, moral theorists need to take into consideration computational modeling methods as another useful tool for normative ethics researches.

## 5. Conclusion

The discussion so far contributes to the situationist debate by proposing a novel agent-based model of the virtue of justice. The results of the simulation experiments based on the model of justice provide an alternative approach to defend the virtue ethics from the situationist objection. Although the argument is not conclusive, it exemplifies how computational modeling methods could be used in the philosophical researches of morality. The analysis of the limitations of the model assumptions indicates at least two promising further research topics: 1. Does the result still hold if we supplement the model of justice with more sophisticated cogni-



tive components that are empirically validated? 2. How can we model other types of virtues, and is it possible to construct a more comprehensive computational model of virtue ethics?

## Works Cited

- Van Zyl, L. *Virtue Ethics: A Contemporary Introduction*[M]. Routledge, 2018, 12.
- Anscombe, G. E. M. ‘Modern Moral Philosophy’[J]. *Philosophy*, 1958, 33(124): 1-19.
- Harman, G. ‘The Nonexistence of Character Traits’[A], *Proceedings of the Aristotelian Society*[C], 100(1): 223-226.
- Doris, J. *Lack of Character: Personality and Moral Behavior*. Cambridge University Press, 2002
- Merritt, M. ‘Virtue Ethics and Situationist Personality Psychology’ [J]. *Ethical Theory and Moral Practice*, 2000, 3(4): 365-383.
- Milgram, S. *Obedience to Authority: An Experimental View*[M]. Harper & Row, 1974.
- Isen, A., and Paula L. 1972. ‘Effect of Feeling Good on Helping: Cookies and Kindness’[J]. *Journal of Personality and Social Psychology*, 1972, 21(3): 384-388.
- Sreenivasan, G. ‘Errors about Errors: Virtue Theory and Trait Attribution’ [J]. *Mind*, 2002, 111(441): 47 – 68.
- Snow, E. N. *Virtue as Social Intelligence: An Empirically Grounded Theory*[M]. Routledge, 2010, 31.
- Miller, C. ‘The Real Challenge to Virtue Ethics from Psychology’[A], Snow, E. N., Trivigno, V. F. (Ed) *The Philosophy and Psychology of Character and Happiness* [C]. Routledge, 2014, 15-34.
- Mascaro, S., Korb, K., Nicholson, A., and Woodberry, O. *Evolving Ethics: The New Science of Good and Evil*[M]. Imprint Academic, 2010.
- Schelling, T. C. ‘Dynamic models of segregation’ [J]. *Journal of Mathematical Sociology*, 1971, 1(2): 143-186.
- Arthur, B. W. ‘Inductive Reasoning and Bounded Rationality’ [J]. *American Economic Review*, 1994, 82(2): 406-411.
- Erev, I., and Roth, E. A. ‘Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria’ [J]. *American Economic Review*, 1998, 88(4): 848-881.
- Smaldino, P. ‘Models Are Stupid, and We Need More of Them’[A], Vallacher, R., Read, S., Andrzej, N.(Ed) *Computational Social Psychology*[C], Routledge: London, 2017, 311-331.