TERG

Discussion Paper No.432

Reinterpreting the General Rules of Morality and the Corruption of Moral Sentiments in The Theory of Moral Sentiments with an Evolutionary Game Model

Takahiko Kan

September 11, 2020

TOHOKU ECONOMICS RESEARCH GROUP

Discussion Paper

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
27-1 KAWAUCHI, AOBA-KU, SENDAI,
980-8576 JAPAN

Reinterpreting the General Rules of Morality and the Corruption of Moral Sentiments in The Theory of Moral Sentiments with an Evolutionary Game Model

Takahiko Kan[1]

1. **Introduction**

Adam Smith is the author of *An Inquiry into the Nature and Causes of the Wealth of Nations*. He is known mainly as a pioneer of political economy. However, he was not only an economist but also a moral philosopher. He published *The Theory of Moral Sentiments* (TMS) in 1759. In TMS, he explained an establishment of a social order based on sympathy between people in a society. Sympathy is sharing of sentiments with others by imaginarily swapping situations with others. People in TMS form the impartial spectator and regulate their conduct to be sympathized by the impartial spectator. The impartial spectator is often considered as an important concept in TMS. However, even if people formed the impartial spectator, this does not mean that they can always regulate their conduct. To regulate their conduct absolutely, people need general rules of morality (GRM). People can establish a social order thanks to GRM.

Some preceding studies have reinterpreted TMS with various research results in contemporary economics. For example, Meardon & Ortmann (1996) reinterprets self-command by using a repeated game theory model. Ashraf *et al.* (2005) indicates that Smith foresaw some research findings of behavioral economics. Tajima (2007) reinterprets TMS from a perspective of institutional economics. Bréban (2012) formularizes a behavior of people in TMS with a utility function, and compares this function with utility functions in behavioral economics. Khalil (2017) reinterprets TMS from a perspective of rational choice theory. These reinterpretations have shed light on modern significance of TMS.

However, there is room for reinterpreting important concepts in TMS with research results in contemporary economics. Following the preceding studies, this paper reinterprets the GRM formation process and the

[1] Graduate School of Economics and Management, Tohoku University, Sendai, Japan. Mail: right1414@yahoo.co.jp

corruption of moral sentiments (CMS) by using a replicator dynamics model, which is a basic model of evolutionary game theory. GRM are the social norms in TMS that concern what is fit and proper either to be done or to be avoided. In TMS, people form GRM through interactions with others. They continually observe conduct of others, and this can lead them to form certain GRM. This paper interprets this observation process as a trial-and-error learning process. To formularize this process, this paper uses a replicator dynamics model.

The results of the model clarify the character of sympathy in the CRM. The more sympathetic players exist in a player set, the more corrupted situation is likely to be realized. This result mathematically supports an interpretation in preceding studies (Brown 1994, Griswold 1999) that sympathy involves risk that CMS is progressing.

The paper is organized as follows. In the section 2, we briefly describe GRM and the CMS. In the section 3, we construct a model of replicator dynamics. In the section 4, we discuss the results of the model. In the last section, we conclude this paper.

2. The Theory of Moral Sentiments

2.1 General Rules of Morality

People in TMS form the impartial spectator and regulate their conduct to be sympathized by the impartial spectator. The impartial spectator is often considered as an important concept in TMS. However, even if people formed the impartial spectator, this does not mean that they can always regulate their conduct. This is because "so partial are the views of mankind with regard to the propriety of their own conduct, both at the time of action and after it; and so difficult is it for them to view it in the light in which any indifferent spectator would consider it" (TMS. III.iv.4). Smith called this kind of problematic situation "self-deceit" (TMS.III.iv.4). According to Smith, to cope with self-deceit, Nature gave people GRM. GRM are the social norms in TMS concerning what is fit and proper either to be done or to be avoided. "General rules of conduct, when they have been fixed in our mind by habitual reflection, are of great use in correcting the misrepresentations of self-love concerning what is fit and proper to be done in our particular situation" (TMS. III.iv.12). GRM are necessary to maintain social orders. "The regard to those general rules of conduct, is … the only principle by which the bulk of mankind are capable of directing their actions" (TMS. III. v.1). Smith referred to cases that people do not have to follow GRM strictly. However, he held that people should strictly follow GRM of virtue of justice, which is essential for the establishment of the social order in TMS. Virtue of justice requires not doing "real and positive hurt to some particular persons, from motives, which are naturally disapproved of" (TMS.II.ii.1.5). "If it [justice] is removed, the great, the immense fabric of human society… must in a moment crumble into atoms" (TMS.II.ii.3.4).

GRM about actions to be avoided are formed thorough the following process. People's sentiments are shocked from some immoral action of others, with which the impartial spectator has no sympathy. People feel this kind of action unseemly. When people hear others express the same feeling as them, "this still further confirms, and even exasperates our [their] natural sense of their deformity" (TMS.III.4.7). After repeatedly hearing others express the same feeling as them, people "resolve never to be guilty of the like, nor ever, upon any account, to render ourselves in this manner the objects of universal disapprobation" (TMS.III.4.7). Then, people form a GRM about an action to be avoided. GRM about actions to be done are formed through a similar process as the GRM about actions to be avoided.

2.2 The Corruption of Moral Sentiments

Smith added a chapter about corruption of moral sentiments (CMS) in the sixth edition of TMS. According to Smith, disposition to "admire, and almost to worship, the rich and the powerful, and to despise, or, at least, to neglect persons of poor and mean condition, though necessary both to establish and to maintain the distinction of ranks and the order of society, is, at the same time, the great and most universal cause of the corruption of our moral sentiments" (TMS.I.iii.3.1). He distinguished two moral principles: the road to virtue and the road to fortune. People in the road to virtue try to be respected "by the study of wisdom and the practice of virtue", whereas people in the road to fortune do so "by the acquisition of wealth and greatness". The road to fortune can involve vices and follies, which contradict the road to virtue.

In explaining why the CMS occurs, Smith distinguished two different ranks of life: middling–inferior stations of life and superior stations of life (Brown 1994, 35). This distinction characterizes the CMS (Hill 2006, 650). In superior stations, the road to fortune causes the CMS, because, in superior stations, "success and preferment depend, not upon the esteem of intelligent and well-informed equals, but upon the fanciful and foolish favour of ignorant, presumptuous, and proud superiors" (TMS.I.iii.3.6). In superior stations, the impartial spectator is too weak to prevent people from corrupting (Hill 2006, 650–651).

In middling–inferior stations, the road to fortune basically causes no corruption, but people in these stations admire and imitate people in superior stations. This happens due to a character of sympathy with joy (Tajima 2007, 591). People in middling–inferior stations imitate people in superior stations, and this cause the CMS in middling–inferior stations (TMS.I.iii.3.7). Superior stations popularize not only a fashion but also vices and follies.

Some authors consider that the CMS was the dominant Smith's motive for publishing the sixth edition (Dickey 1986, 608;32 Evensky 1989, 131; Dwyer 2005, 684; Forman-Barzilai 2010, 97).

3 **Formularizing the formation process of GRM**

3.1 *The model*

As we briefly explained above, GRM are "formed, by finding from experience, that all actions of a certain kind, or circumstanced in a certain manner, are approved or disapproved of" (TMS.III.4.8). This formation process of GRM can be regarded as a process of trial-and-error. Trial-and-error learning is a kind of empirical learning, through which people evaluate their choice based on its consequence and modify their way of choice. In the formation process, people evaluate justifiability of their feeling based on whether others express the same feeling as them or not. When they know others express the same opinion as them, this justifies their opinion. People evaluate justifiability of their feeling (choice) based on whether others express the same opinion (consequence).

To formularize the learning process, this paper uses a replicator dynamics model with two strategies. Players have their mixed strategies on the two strategies. The set of players is $I = \{1,2,\dots,n\}$, where $n = 2m(m \in \mathbb{N})$. The set of pure strategies of player i is $S_i = \{A, B\}$. We consider strategy A as *virtue strategy*. We consider strategy B as *fortune strategy*. These two strategies derive from the road to virtue and the road to fortune in the CMS (TMS. I.iii.3.1).

The mixed strategy of player $i \in I$ is $x_i = (x_{iA}, x_{iB}) \in [0,1]^2$. The profile of mixed strategies of players is $x = (x_1, x_2, .., x_n) \in [0,1]^{2\times n}$.

In the model, choosing a strategy means that the player considers he should choose this strategy. Justifiability of each strategy is reflected in mixed strategies. We assume probabilistic choice of players because, before GRM is formed, people in TMS probabilistically choose what they should do. Before GRM is formed, people sometimes fail to choose what the impartial spectator suggests because they suffer from "self-deceit" (TMS. III.iv.4). Once GRM is formed, they can surely choose what the impartial spectator suggests.

Trial-and-error learning processes have been formularized by using some evolutionary game models.[2] This paper uses a simple replicator dynamics model among evolutionary game models, because other models involve assumptions which cannot be supported by sentences in TMS.

Each player is randomly matched with one player and then observes the choice of the opponent. This matching with observation is continuously repeated.

Each i's x_{iA} changes depending on their payoffs of each game. Pure strategies of player i is $s_i \in S_i$. The profile of pure strategies is $s = (s_1, \dots, s_n)$, and $s \in S = \times_{i=1}^{n} S_i$. For any $s \in S$, payoff of player i is defined with pure strategy payoff function $\pi_i : S \to \mathbb{R}$.

When mixed strategy profile $x \in [0,1]^{2\times n}$ is given, $s = (s_1, \dots, s_n) \in S$ has the probability $x(s) = \prod_{i=1}^{n} x_{is_i} \in [0,1]$.

Let $u_i : [0,1]^{2\times \mathrm{n}} \to \mathbb{R}$ be

$$u_i(x) = \sum_{s \in S} x(s)\pi_i(s).$$

Function u_i is expected payoff of player i when $x \in [0,1]^{2\times n}$ is played.

[2] See, for example, Roth & Erev(1995) and Erev & Roth(1998).

Let $e_i^{s_i}$ denote the mixed strategy of i when i chooses pure strategy s_i. Expected payoff of player i when i chooses $e_i^{s_i}$ is $u_i(x_1, \dots, e_i^{s_i}, \dots x_n)$.

To know the behavior of x, knowing the behavior of $x_{\mathrm{A}} = (x_{1A}, \dots, x_{nA}) \in [0,1]^n$ is sufficient. Let the dynamics of x_{iA} be

$$\dot{x}_{iA} = \left[u_i(x_1, \dots, e_i^A, \dots x_n) - u_i(x)\right] x_{iA},$$

where $u_i(x)$ is the expected payoff of i when x is played, and $u_i(x) = x_{iA} \cdot u_i(x_1, \dots, e_i^A, \dots x_n) + x_{iB} \cdot u_i(x_1, \dots, e_i^B, \dots x_n)$ holds. In the dynamics, the higher expected payoff of a strategy is, the higher the probability of choosing the strategy becomes.

The payoff matrix of i is described as follows.

player i／ opponent	A	B
A	a_{AA}^i	a_{AB}^i
B	a_{BA}^i	a_{BB}^i

Table Payoff of player i

Let us define $\alpha_i \equiv a_{AA}^i - a_{BA}^i$ and $\beta_i \equiv a_{AB}^i - a_{BB}^i$. The simultaneous differential equations of the dynamics are the following. To derive the equations, we used $u_i(x_1, \dots, e_i^A, \dots x_n) = a_{AA}^i(\sum_{j\neq i} x_{jA}/n - 1) + a_{AB}^i\left(1 - (\sum_{j\neq i} x_{jA}/n - 1)\right)$, and $u_i(x_1, \dots, e_i^B, \dots x_n) = a_{BA}^i(\sum_{j\neq i} x_{jA}/n - 1) + a_{BB}^i\left(1 - (\sum_{j\neq i} x_{jA}/n - 1)\right)$.

$$\dot{x}_{1A} = \left((\alpha_1 - \beta_1)\left({}^{\sum_{j\neq 1} x_{jA}}\!/_{n-1}\right) + \beta_1 \right) x_{1A}(1 - x_{1A}),$$

$$\vdots$$

$$\dot{x}_{iA} = \left((\alpha_i - \beta_i)\left({}^{\sum_{j\neq i} x_{jA}}\!/_{n-1}\right) + \beta_i \right) x_{iA}(1 - x_{iA}),$$

$$\vdots$$

$$\dot{x}_{nA} = \left((\alpha_n - \beta_n)(\sum_{j\neq n} x_{jA} / _{n-1}) + \beta_n \right) x_{nA}(1 - x_{nA})。$$

From these equations, all edge points are equilibrium points. In this dynamics, all interior points are not asymptotically stable. We can proof this result by using the proposition 6.6 in Weibull(1997). To consider the asymptotically stable points, we only need to check edge points.

This paper assumes that there are three types of players. These types are defined based on payoff matrices.

Player i is *sympathy type* if $\alpha_i > 0$, $\beta_i < 0$, $a^i_{AB} < a^i_{AA}$, $a^i_{BA} < a^i_{BB}$. Sympathy type players get higher payoff when they choose the same strategy as their opponents.

Player i is *virtue type* if $\alpha_i > 0$, $\beta_i > 0$. Virtue type players get higher payoff when they choose strategy A than when they choose strategy B. This means $\dot{x}_{iA} > 0$ holds except at equilibrium points.

Player i is *fortune type* if $\alpha_i < 0$, $\beta_i < 0$. Fortune type players get higher payoff when they choose strategy B than when they choose strategy A. This means $\dot{x}_{iA} < 0$ holds except at equilibrium points.

3.2 Seven Cases of Games

There are seven cases of games because we assume three types of players. Results of the seven cases are shown in the following. This paper provides supplementary information in the appendix.

Case 1 All players are sympathy type

Both $(0,0,\ldots,0)$ and $(1,1,\ldots,1)$ are asymptotically stable. Asymptotical stability of the other edge points depends on each player's α_i and β_i.

Case 2 All players are fortune type

$(0,0,\ldots,0)$ is the unique asymptotically stable point because $\dot{x}_{iA} < 0$ holds for any i except at equilibrium points.

Case 3　All players are virtue type

(1,1, … ,1) is the unique asymptotically stable point because $\dot{x}_{iA} > 0$ holds for any i except at equilibrium points.

Case 4　Sympathy type and fortune type

(0,0, … ,0) is asymptotically stable, but (1,1, … ,1) is not asymptotically stable. Asymptotical stability of other edge points depends on each player's α_i and β_i.

Case 5　Sympathy type and virtue type

(1,1, … ,1) is asymptotically stable, but (0,0, … ,0) is not asymptotically stable. Asymptotical stability of other edge points depends on each player's α_i and β_i.

Case 6　Virtue type and fortune type

Without loss of generality, assume player 1 to player p are fortune type. The asymptotically stable point is unique, and, at this point, the first element to the pth elements are all 0, and the rest elements are all 1. This is because, except at equilibrium points, $\dot{x}_{iA} < 0$ holds for any fortune type player, and $\dot{x}_{iA} > 0$ holds for any virtue type player.

Case 7　Three types

Asymptotically stable points vary depending on each player's α_i and β_i.

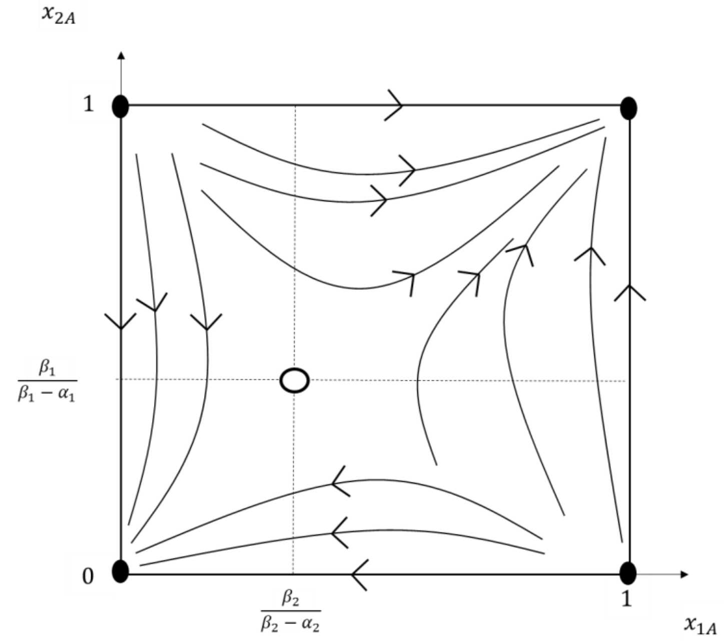
x_{2A}
1
$\frac{\beta_1}{\beta_1-\alpha_1}$
0
$\frac{\beta_2}{\beta_2-\alpha_2}$
1
x_{1A}

Figure　Two sympathy type players game

4 **Discussion**

Smith considered the road to fortune causes the CMS. We can consider that the existence of fortune type characterizes the CMS. That is why this paper clarifies what the existence of fortune type means. This paper compares case 3 and 6 first. Then this paper compares case 1 and 4, and case 5 and 7.

We define a degree of corruption as the number of players choosing fortune strategy. By this definition, the most corrupted point is $(0,0,\dots,0)$. In contrast, the least corrupted point is $(1,1,\dots,1)$, where we can consider a GRM is formed.

Let us assume a change from case 3 to 6. Adding fortune type players increases the degree of corruption at the asymptotically stable point by the number of the fortune type. When the number of the fortune type is less, the degree of corruption is also less.

Next, let us assume a change from case 1 to 4. In case 1, both $(0,\dots,0)$ and $(1,\dots,1)$ are asymptotically stable regardless of each player's α_i and β_i. In case 4, $(1,\dots,1)$ becomes not asymptotically stable, and $(0,\dots,0)$ becomes the unique point that is asymptotically stable regardless of each player's α_i and β_i. Adding merely one fortune type player makes the most corrupted point, $(0,\dots,0)$, be the unique point.

Last, let us assume a change from case 5 to 7. In case 5, $(1,1,\dots,1)$, the least corrupted point, is asymptotically stable regardless of each player's α_i and β_i. This never holds once adding merely one fortune type player.

As shown above, adding fortune type players increases the degree of corruption of asymptotically stable points. In case 1 and 5, such an addition can increase the degree of corruption more than by the number of the fortune type. We can consider such increase of the degree of corruption is more severe in case 1 (to 4) than in case 5 (to 7). This is because, only in case 4, the most corrupted point $(0,\dots,0)$ is the unique asymptotically stable point. Thus, the more sympathy type players exist, the more corrupted situation fortune type players cause.

case	Change
1 to 4	$(1,1,\dots,1)$ becomes not asymptotically stable. $(0,\dots,0)$ becomes the unique asymptotically stable point.
3 to 6	The degree of corruption of asymptotically stable points increases by the number of the fortune type players.
5 to 7	$(1,1,\dots,1)$ becomes not asymptotically stable.

Table　Adding fortune type players

5 **Concluding Remarks**

This paper provides a formularization of corruption of moral sentiments (CMS), which is consistent with the formation process of general rules of morality (GRM). The model reveals that the more sympathy type players exist, the more corrupted situation fortune type players cause.

The result is consistent with an interpretation in preceding studies (Brown 1994; Griswold 1999) that sympathy involves risk that CMS is progressing. Sympathy type players are consistent with sympathy as its name suggests. This paper mathematically supports the interpretation in preceding studies.

Appendix

Let us consider the Jacobian matrix of the simultaneous differential equations of the dynamics. A necessary and sufficient condition for equilibrium points to be asymptotically stable is that the Jacobian is stable. In the following, we consider only equilibrium edge points. This is because all interior points are not asymptotically stable in the dynamics. We can proof this result by using the proposition 6.6 in Weibull(1997).

Diagonal component of the Jacobian at ith row is

$$\frac{\partial \dot{x}_{iA}}{\partial x_{iA}} = \left((\alpha_i - \beta_i)\left(\sum_{j \neq i} x_{jA} \Big/ n-1\right) + \beta_i \right)(1 - 2x_{iA}).$$

Non-diagonal component of the Jacobian at ith row is

$$\frac{\partial \dot{x}_{iA}}{\partial x_{jA}} = \frac{\alpha_i - \beta_i}{n-1} x_{iA}(1 - x_{iA}).$$

At equilibrium edge points, non-diagonal components of the Jacobian are zero. Thus, a necessary and sufficient condition for equilibrium edge points to be asymptotically stable is that all diagonal components of the Jacobian are negative.

We can get the results of case 1 to 7 by using the above information. Let us consider case 1 for instance.

Case 1 All players are sympathy type

$(0,0,\dots,0)$ *and* $(1,1,\dots,1)$ *are asymptotically stable. Asymptotical stability of the other edge points depends on each player's* α_i *and* β_i*.*

At $(0,0,\dots,0)$ and $(1,1,\dots,1)$, all diagonal components of the Jacobian are negative. Thus, $(0,0,\dots,0)$ and $(1,1,\dots,1)$ are asymptotically stable. At other edge points, the sign of diagonal components can be positive depending on each player's α_i and β_i. For example, at $(0,0,\dots,0,1)$, the diagonal component of the Jacobian at first row is $(\alpha_1 - \beta_1)\left(1 \big/ n-1\right) + \beta_1$, which can be positive with some $\alpha_1$and β_1 because player 1 is sympathy type.

References

- Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives*, 19(3), 131-145.
- Bréban, L. (2012). Sensitivity to prosperity and adversity: What would a Smithian function of happiness look like?. *The European Journal of the History of Economic Thought*, 19(4), 551-586.
- Brown, V. (1994). *Adam Smith's discourse: canonicity, commerce and conscience*. London: Routledge.
- Dickey, L. (1986). Historicizing the" Adam Smith Problem": Conceptual, historiographical, and textual issues. *The Journal of Modern History*, 58(3), 580-609.
- Dwyer, J. (2005). Ethics and Economics: Bridging Adam Smith's Theory of Moral Sentiments and Wealth of Nations. *Journal of British Studies*, 44(4), 662-687.
- Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American economic review*, 88(4), 848-881.
- Evensky, J. (1989). The evolution of Adam Smith's views on political economy. *History of political economy*, 21(1), 123-145.
- Forman-Barzilai, F. (2010). *Adam Smith and the circles of sympathy: cosmopolitanism and moral theory*, Cambridge: Cambridge University Press.
- Griswold Jr, C. L. (1999). *Adam Smith and the virtues of enlightenment*. Cambridge: Cambridge University Press.
- Hill, L. (2006). Adam Smith and the theme of corruption. *The Review of Politics*, 68(4), 636-662.
- Khalil, E. L. (2017). Socialized view of man vs. rational choice theory: What does smith's sympathy have to say?. *Journal of Economic Behavior & Organization*, 143, 223-240.
- Meardon, S. J., & Ortmann, A. (1996). SELF-COMMAND IN ADAM SMITH'S THEORY OF MORAL SENTIMENTS A GAME-THEORETIC REINTERPRETATION. *Rationality and Society*, 8(1), 57-80.
- Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, 8(1), 164-212.
- Smith, A. (1976a). *An inquiry into the nature and causes of the wealth of*

nations. Ed. R.H. Campbell & A.S. Skinner, Oxford: Oxford University Press.

- Smith, A. (1976b). *The theory of moral sentiments*. Ed. D.D. Raphael & A.L. Macfie, Oxford: Oxford University Press.
- Tajima, K. (2007). The theory of institutions and collective action in Adam Smith's Theory of Moral Sentiments. *The Journal of Socio-Economics*, 36(4), 578-594.
- Weibull, J. W. (1997). *Evolutionary game theory*. Cambridge, MA: MIT Press.