Moral Guilt : An Agent-Based Model Analysis

Benoit Gaudou, Emiliano Lorini, Eunate Mayor

Moral Guilt: An Agent-Based Model Analysis

Benoit Gaudou[1,2], Emiliano Lorini[1], and Eunate Mayor[1]

[1] UMR 5505 CNRS IRIT, Toulouse, France
[2] University of Toulouse, Toulouse, France
benoit.gaudou@ut-capitole.fr, emiliano.lorini@irit.fr,
eunate.mayor@irit.fr

Abstract. In this article we analyze the influence of a concrete moral emotion (i.e. moral guilt) on strategic decision making. We present a normal form Prisoner's Dilemma with a moral component. We assume that agents evaluate the game's outcomes with respect to their ideality degree (*i.e.* how much a given outcome conforms to the player's moral values), based on two proposed notions on ethical preferences: Harsanyi's and Rawls'. Based on such game, we construct and agent-based model of moral guilt, where the intensity of an agent's guilt feeling plays a determinant role in her course of action. Results for both constructions of ideality are analyzed.

1 Introduction

Few aspects of human evolution have been more controversial that the explanation of human ethics and morality. Especially when natural selection theory reached its peak, a question started to be posed more and more often: is cooperation compatible with this phenomena? Or is it so that only selfish behavior can survive under such circumstances? And what about cooperation and other-regarding behavior?

According to Dawkins, all factors that lead to the evolving of instincts that favor other-regarding behavior can be summarized into four main types: "We now have four good Darwinian reasons for individuals to be altruistic, generous or 'moral' towards each other. First, there is the special case of genetic kinship. Second, there is reciprocation: the repayment of favors given, and the giving of favors in 'anticipation' of payback. [...] [T]hird, the Darwinian benefit of acquiring a reputation for generosity and kindness. And fourth, [...] there is the particular additional benefit of conspicuous generosity as a way of buying unfakeably authentic advertising."[11]. For Dawkins, through most of our prehistory, humans lived under conditions that would have strongly favored the evolution of other-regarding tendencies. The social side of our species, motivates that, whether kin or not, individuals would tend to meet again and again throughout their lives, favoring other-regarding behaviors.

Economic theories based in the self-regarding assumption have stated that, except for sacrifice on behalf of others (what we call 'altruism'), the rest is just long-run material self-interest, such theories abstract from reciprocity and other non-self-regarding motives which can guide individuals' behavior. Thus, although cooperation among purely self-regarding agents in indefinitely repeated games with sufficiently low discount rate is a widely accepted theoretical result, this narrow interpretation challenges

observations of our everyday life. Indeed, there is compelling evidence that individuals adhere to norms of fairness in both experimental and real world situations that are non-repeated or infrequently repeated [2]. People do repay gifts and take revenge in interactions with complete strangers, even in those cases where it is costly for them and yields to neither present nor future material rewards[1].

Moreover, human beings act cooperatively, obey and enforce norms of fairness, even against their self-interest, and the volume of experimental evidence supporting these and other facts that separate us from the selfishness assumption continues to grow (see, for example [12]). The assumption that individuals are self-regarding is in strong conflict with daily observed preferences. *First,* because agents not only care about the outcomes of their economic interactions, but also about the process through which the results are attained. *Second,* because in their decisions agents do not solely consider what they *personally* gain and lose through an interaction. Violating a fairness norm has emotional consequences that enter negatively in the agent's utility function [7]. In addition,we can say that adherence to norms of fairness is underwritten by emotions, and not merely by the expected gain from the repeated interaction [7] (the so-called 'prosocial emotions', such as shame, guilt, empathy or remorse, all of which involve feelings of discomfort at doing something that appears wrong according to one's own values and/or those of other agents whose opinions one values). Furthermore, social scientists (*e.g.*, [6]) have defended the idea that there exist innate moral principles in humans such as fairness which are the product of biological evolution.

In this article, we test the hypothesis that agents have fairness as a moral value, and that the transgression of this moral value of fairness triggers guilt feelings in them. In order to measure the influence of *moral guilt* on the agents' decision-making process, we present an agent-based model of the Prisoner's Dilemma, where the intensity of an agent's guilt feeling plays a determinant role in her course of action. The paper is organized as follows. In Section 2.1 we present an overview of the concept of guilt and the analysis of our game-theoretic model of moral guilt (based on two proposed notions of moral values: Harsanyi's and Rawls') and its influence on strategic decision making. In Section 3 we describe the agent-based model and its implementation. Finally, in Section 3.2 we present some preliminary results and, in Section 4, our ideas for future work.

2 Moral guilt

While there exist many contrasting theories explain the discrepancy between pure intentional decision and moral behavior, most of them highlight the variability of individual behavior depending on the situational context and group belonging[2].

[1] Cf. J. Mansbridge's monograph [20], where several social scientists from different disciplines argue that individuals have motives for action that go well beyond their egoistic desires and pure rational calculations. People, they suggest, are influenced by feelings of solidarity, altruism and concern for others and their well-being.

[2] A well-integrated model of the ways in which attitudes, norms, and perceived control feed into behavioral intentions and subsequent behavior is proposed by Ajzen's theory of planned behavior [1].

Furthermore, adherence to moral standards and social norms is underwritten by emotions, the so-called, *prosocial emotions*, such as empathy, shame, guilt, pride or regret. The influence of these type of emotions in the agents' behavior is two-fold [7]: *on the one hand,* they have emotional consequences that affect negatively the agent's preference function; and *on the other hand,* they induce the agent to act in ways that increase the average payoff to other members of the group to whom she belongs.

There are two main trends in guilt literature. *On the one hand,* part of the scholarship considers guilt a belief-based emotion, what is referred to as *'interpersonal guilt'*. The fact that an agent's utility is 'belief-based', in the sense of second-order beliefs (*i.e.* beliefs about other agents' beliefs) is also well-accepted in the literature. The latter has been explained in two ways. *First,* according to the 'social esteem model', where agents care about what others think about them, and thus it represents an element in their utility function (see notably [5]). *Second,* by means of the 'guilt aversion model', were agents care about what others expect of them; that is, agents feel guilty for "hurting their partners [...] and for failing to live up to their expectations, [which motivated them to] alter their behavior [to avoid guilt]."[4] (see also [3] and [9]). *On the other hand,* theories of *'moral guilt'* define guilt as a 'self-conscious' emotion, triggered by the violation of one's moral standards and internalized (social) norms. In this paper, we present a model of guilt in this latter sense.

2.1 The model

We present a game-theoretic analysis of normative guilt and of its influence on strategic decision making. The intensity of a player's guilt feeling is defined as the difference between the degree of ideality of the *actual* state and the degree of ideality of the *counterfactual* state that could have been achieved had the player chosen a different action. The model assumes a player has two different motivational systems: an endogenous motivational system determined by the player's desires and an exogenous motivational system determined by the player's moral values. Moral values, and more generally moral attitudes (ideals, standards, etc.), originate from an agent's capability of discerning what from his point of view is (morally) *good* from what is (morally) *bad*. If an agent has a certain moral value, then he thinks that its realization ought to be promoted because it is *good* in itself. A similar distinction has also been made by philosophers and by social scientists. For instance, Searle [24] has recently proposed a theory of how an agent may want something without desiring it and on the problem of reasons for acting based on moral values and independent from desires. In his theory of morality [17, 16], Harsanyi distinguishes a person's *ethical preferences* from her *personal preferences* and argues that a moral choice is a choice that is based on ethical preferences.

2.2 Guilt-dependent utility

Let us first introduce the notion of normal form game.

Definition 1 (Normal form game). *A normal form game is a tuple* $\Gamma = (N, \{S_i\}_{i \in N}, \{U_i\}_{i \in N})$ *, where:*

- $N = \{1, \ldots, n\}$ *is a set of players;*

- S_i *is player i's set of strategies;*
- $U_i : \prod_{i \in N} S_i \longrightarrow \mathbb{R}$ *is agent i's personal utility function mapping every strategy profile in* $\prod_{i \in N} S_i$ *to a real number (*i.e.*, personal utility of the strategy profile for player i).*

Let $2^{Agt*} = 2^N \setminus \{\emptyset\}$ be the set of all non-empty sets of players (*alias* coalitions). For notational convenience we write $-i$ instead of $N \setminus \{i\}$. For every $J \in 2^{Agt*}$, we define the set of strategies for the coalition J to be $S_J = \prod_{i \in J} S_i$. Elements of S_J are denoted by $s_J, s'_J, \ldots$ For notational convenience, we write S instead of S_N and we denote elements of S by $s, s', \ldots$ Every strategy s_J of coalition J can be seen as a tuple $(s_i)_{i \in J}$ where player i chooses the individual strategy $s_i \in S_i$.

The following definition extends the definition of normal form game with a *moral* component. Namely we assume that players in a game also evaluates outcomes with respect to their ideality degree, *i.e.*, how much a given outcome conforms to the player's moral values.

Definition 2 (Normal form game with moral values). *A normal form game with moral values is a tuple* $\Gamma^+ = (N, \{S_i\}_{i \in N}, \{U_i\}_{i \in N}, \{I_i\}_{i \in N})$ *where:*

- $(N, \{S_i\}_{i \in N}, \{U_i\}_{i \in N})$ *is a normal form game;*
- $I_i : \prod_{i \in N} S_i \longrightarrow \mathbb{R}$ *is agent i's ideality function mapping every strategy profile in* $\prod_{i \in N} S_i$ *to a real number (*i.e.*, the ideality of the strategy profile for player i).*

The preceding notion of ideality corresponds to Harsanyi's notion of ethical preference.

We define guilt as the emotion which arises from the comparison between the ideality of the current situation and the ideality of a counterfactual situation that could have been achieved had the player chosen a different action. In particular, intensity of guilt feeling is defined as the difference between the ideality of the current state and the ideality of the best alternative state that could have been achieved had the player chosen a different action.

Definition 3 (Guilt). *Given a normal form game with moral values* $\Gamma^+ = (N, \{S_i\}_{i \in N}, \{U_i\}_{i \in N}, \{I_i\}_{i \in N})$ *the guilt player i will experience after the strategy profile s is played, denoted by Guilt(i,s), is defined as follows:*

$$Guilt(i,s) = I_i(s) - \max_{a_i \in S_i} I_i(a_i, s_{-i}) \tag{1}$$

We assume that guilt affects the utility function of a certain player depending on the player's degree of guilt aversion. More precisely, the higher the influence of guilt on the utility of a given decision option, the more guilt averse the player. The extent to which a player's utility is affected by his guilt feeling is called *degree of guilt aversion*. The following definition describes how a player's utility function is transformed depending on the player's guilt and on the player's degree of guilt aversion.

Definition 4 (Guilt-dependent utility). *Given a normal form game with moral values* $\Gamma^+ = (N, \{S_i\}_{i \in N}, \{U_i\}_{i \in N}, \{I_i\}_{i \in N})$ *the guilt-dependent utility of the strategy profile s for agent i is defined as follows:*

$$U_i^*(s) = U_i(s) + \delta_i(Guilt(i,s)) \tag{2}$$

where δ_i *is a nondecreasing function* $\delta_i : \mathbb{R} \longrightarrow \mathbb{R}$ *such that* $\delta_i(0) = 0$.

The previous definition of guilt-dependent utility is related with the definition of regret-dependent utility proposed in regret theory [18, 19, 15]. Specifically, similarly to Loomes & Sugden's regret theory, we assume that computation of emotion-dependent utility consists in adding to player i's personal utility the value $\delta_i(Emotion(i,s))$ which measures the intensity of player i's current emotion.[3] There are several possible instantiations of the function $\delta_i(Guilt(i,s))$. For example, it might be defined as follows:

$$\delta_i(Guilt(i,s)) = c_i \times Guilt(i,s) \tag{3}$$

where $c_i \in \mathbb{R}^+ = \{x \in \mathbb{R} | x \geq 0\}$ is a constant measuring player i's degree of guilt aversion.

2.3 Grounding moral values on personal utilities

In the preceding definition of normal form game with moral values a player i's utility function U_i and ideality function I_i are taken as independent. Harsanyi's theory of morality provides support for an utilitarian interpretation of moral motivation which allows us to reduce a player i's ideality function I_i to the utility functions of all players [17, 16]. Specifically, Harsanyi argues that an agent's moral motivation coincides with the goal of maximizing the collective utility represented by the weighted sum of the individual utilities.

Definition 5 (Normal form game with moral values based on Harsanyi's view). *A normal form game with moral values* $\Gamma^+ = (N, \{S_i\}_{i\in N}, \{U_i\}_{i\in N}, \{I_i\}_{i\in N})$ *is based on Harsanyi's view of morality if and only if for all* $i \in N$*:*

$$I_i(s) = \sum_{j\in N} k_{i,j} \times U_j(s) \tag{4}$$

for some $k_{i,1}, \ldots, k_{i,n} \in [0, 1]$.

The parameter $k_{i,j}$ in the previous equation can be conceived as the agent i's *degree of empathy* towards agent j. This means that the higher the degree of empathy of agent i towards agent j, the higher the influence of agent j's personal utility on the degree of ideality of a given alternative for agent i. In certain situations, it is reasonable to suppose that an agent has a maximal degree of empathy towards all agents, *i.e.*, $k_{i,j} = 1$ for all $i, j \in N$. Under this assumption, the previous equation can be simplified as follows:

$$I_i(s) = \sum_{j\in N} U_j(s) \tag{5}$$

An alternative to Harsanyi's utilitarian view of morality is Rawls' view [22]. In response to Harsanyi, Rawls proposed the *maximin* criterion of making the least happy agent as happy as possible: for all alternatives s and s', if the level of well-being in the worst-off position is strictly higher in s than in s', then s is better than s'. According to

[3] On the ground of empirical evidence, Loomes & Sugden also suppose that the function δ_i should be convex. To keep our model simpler, we do not make this assumption here.

this well-known criterion of distributive justice, a fair society should be organized so as to admit economic inequalities to the extent that they are beneficial to the less advantaged agents.[4] Following Rawls' interpretation, an agent's moral motivation should coincide with the goal of maximizing the collective utility represented by the individual utility of the less advantaged agent.

Definition 6 (Normal form game with moral values based on Rawls' view). *A normal form game with moral values* $\Gamma^+ = (N, \{S_i\}_{i\in N}, \{U_i\}_{i\in N}, \{I_i\}_{i\in N})$ *is based on Rawls' view of morality if and only if for all* $i \in N$*:*

$$I_i(s) = \min_{j\in N} U_j(s) \tag{6}$$

3 Simulation

We have developed an agent-base model in order to test the various hypothesis on the model. In this article, we mainly investigate the influence of the way to compute ideality and the influence of the guilt aversion on agent behaviors. To this purpose, we have chosen the standard Prisoner's Dilemna [21] as frame of the interactions between agents.

The model is implemented using the GAMA platform[5] [26], an open-source generic agent-based modeling and simulation platform. It provides an intuitive modeling language with high-level primitives to define agents and their environment. GAMA has been used successfully to develop a large spectrum of models, from simple and abstract models (as the one presented here) to large-scale models (including lot of different kinds of agents and needing a huge amount of data).

3.1 Model description

The following paragraphs describe in greater details the computational implementation of the model described in section 2.1.

Global variables and parameters

The game that agents will play is the same for all agents, it (and in particular the payoff matrix) is thus a global variable. It is a standard Prisoner's Dilemma whose payoffs are the following: R = 2, T = 3, S = 0 and P = 1[6]. Being a standard Prisoner's Dilemma type of game, each agent has two possible strategies: cooperate (C) or defect (D). In addition, as we will see, we shall modify the ideality computation method (*i.e.*

[4] It has to be noted that Rawls' theory of justice is specified in terms of justice over primary goods. Rawls' list of primary goods includes for instance basic liberties and rights, freedom of movement and free choice of occupation, income and wealth, the social bases of self-respect. This difference is however beyond the scope of the present article. See [25] for a discussion on this issue.

[5] http://code.google.com/p/gama-platform/

[6] Although for this first version of the article, we employ the standard payoffs mentioned, the four payoff values might be modified, as they are parameters of the simulation.

Harsanyi's and Rawls' measures) in order to test their influence on the simulation results. The number of time-steps of the simulation might also be changed, allowing us to analyze the influence of the learning process on the results. Finally, the maximum guilt aversion level and the discretization step of the guilt aversion may also be altered.

Agents

The model is composed of one unique kind (or *species*) of agents named 'people'. Each agent is characterized by a guilt aversion level (*guiltAversion*), a positive float number lower or equal to the global parameter (*guiltAversionInitMax*), and an *history* of previous interactions. Each agent *i*'s *history* is a complex structure (a mapping function) associating each other agent *j* already met to a list containing: (1) the number of interactions between both agents, (2) the number of interactions in which *j* chose to cooperate with *i*, and (3) the overall payoff won by *i* from such interactions with agent *j*. As we will see in the following paragraph, the two first elements of the list are taken into account in the computation of the expected utility, whereas the last one is only an indicator of the 'quality' of the interaction between both agents. Derived from the expected utility obtained from the combination of these two first elements, each agent will compute a guilt dependent utility matrix (containing a modified utility U^*value) from the game utility matrix. It is also important to note that agents are not aware of the guilt aversion level of their interaction partner. They thus make their decision only depending on other agents behavior (their moves).

Learning: fictitious play

In order to explain Nash equilibrium (and selection among various Nash equilibria), game theorist have traditionally used different kinds of adjustment models (cf. for example [27] or [14]); mainly *replicator dynamics* (*i.e*. the relative prevalence of any strategy has a growth rate proportional to its payoff relative to the average payoff) and *simple belief learning* (*i.e*. players adjust their beliefs as they accumulate experience, and that current beliefs influence the current choice of strategy). As shown empirically by [10], in both symmetric (single population) and two-type population games, "the learning model is slightly better at explaining the single population data and much better at explaining the two population data."

Thus, in our simulation, we use a simple belief learning process known as *'fictitious play'*, or as the 'Brown-Robinson learning process'. The algorithm was introduced by Brown [8] as an algorithm for finding the value of a zero-sum game, and first studied by Robinson [23]. It assumes that players noiselessly best respond to the belief that other players' current actions will be equal to the average of their actions in all earlier periods. Informally, we can describe it as follows. Let us assume two players playing a finite game repeatedly. After arbitrary initial moves in the first round, where each player chooses a single pure strategy; then both players construct sequences of strategies according to the following rule: at each step, a player considers the sequence chosen by the other player, she supposes that the other player will randomize uniformly over that sequence, and she chooses a best response to that mixed strategy. That is, in every round each player plays a myopic pure best response against the empirical strategy distribution of her opponent (a player's sequence is treated as a multi-set of strategies, one of which

is selected uniformly at random). At each time-step, the chosen best-response is added to a player's sequence, and her strategy sequence get extended.

Initialization

The initialization is limited to the creation of the agents and the initialization of the game from simulation parameters. The history, as it has been implemented right now, makes interactions with one agent totally independent from interactions with others: an agent's expected utilities will be computed taking into account only the ratio of cooperation on total interactions with each other agents. We can thus create only one agent per possible guilt aversion level in order to explore all the space of possible interactions given all other parameters. We thus create one agent per guilt aversion level from 0 to a maximal chosen value of the guilt aversion parameter ('*guiltAversionInitMAx*'), for each discretization step (concretely each 0.1, in this case). Although establishing an upper threshold for the guilt aversion parameter might seem arbitrary, in the following section 3.2 we will see that, from a given degree of guilt aversion, results in terms of agents' payoffs do not vary.

Model dynamics

At each simulation time-step we randomly pair agents. For each pair, each agent will first compute the expected utilities associated to each of the possible strategies (C and D), depending on the probability distribution (computed from the history) of their mate's strategies. Note that agents choose blindly the first time the interact and then select, according to the recorded history, the pure best response against the empirical strategy distribution of their opponent (cf. paragraph 3.1); that is, the strategy that maximizes her expected utility. In case both strategies have the same expected utility, agents choose randomly. After choosing, each agent is informed of the strategy the other agent played, and computes her payoffs. Agents update their history: (1) they number of interactions with that given opponent; (2) the number of interactions in which j cooperated, if it is the case; and (3) they payoff won (cf. paragraph 3.1). This payoff represents the real payoff of the game, without taking into account the guilt element; that is, independently of the agent's ideality notion. The simulation is iterated until the limit time-step chosen. The history is then registered to be analyzed afterwards.

3.2 Results: Harsanyi's and Rawls' idealities comparison

In Figures 1 and 2, we illustrate behaviors emerging from the interactions during the game being played. Both axes represent the guilt aversion values from 0 to 10 (with a step of 0.1). Each guilt aversion value represents also one single agent, as we have chosen to create 1 agent per each guilt aversion value. Figures represent thus the behavior that is emerging from the interactions of the agent on the vertical axe with the one on the horizontal axe. Note that we stopped simulations after 5000 steps[7]. In both cases, we launched 20 replications (without observing any variability in the results despite the randomness of the first move).

[7] This number should be big compared to the number of agents, to allow a high ehough number of interactions with all other agents.

Red (resp. green) areas in Figures represent the fact that agents in these interval always play the strategy profile D,D (resp. C,C).

Harsanyi's ideality

We first use Harsanyi's algorithm of ideality in order to compute the agent's modified utility function U^*. For this first simulation, we keep all the experiment parameters at their default values. We can observe results in the following Figure 1, it represents the convergent behavior of agents on vertical axe (i) interactions with agents on the horizontal one (j).

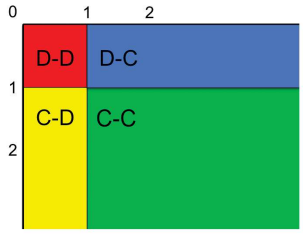

Fig. 1. Harsanyi's ideality results

For Harsanyi's case, there are not very remarkable results. However, we shall note that high guilt aversion does not 'pay off'. On the contrary, less guilt averse agents have a higher average payoff (indeed, guilt averse agents interacting with guilt seeking, get 'cheated on' and their average payoff is the 'sucker' one, that is, zero; whilst their opponent benefits from their defection and obtains the maximum payoff, three).

Furthermore, when implementing Harsanyi's ideality algorithm, the learning process implemented in the agent is not involved into the decision-making process. Indeed, it is easy to see that if an agent has a degree of guilt aversion lower than one, in the game with transformed utility U^* the agent's strategy D strongly dominates the agent's strategy C. Therefore, for every possible probability distribution over the opponent's strategy, D is the strategy which maximizes the agent's expected utility. On the contrary, if an agent has a degree of guilt aversion higher than one, in the game with transformed utility U^* the agent's strategy C strongly dominates the agent's strategy D. Therefore, for every possible probability distribution over the opponent's strategy, C is the strategy which maximizes the agent's expected utility. Hence, an agent with degree of guilt aversion lower than one will always play D, whereas an agent with degree of guilt aversion higher than one will always play C. Agents with degree of guilt aversion equal to one are exactly in the co-joint point where the expected utilities of the two strategies C and D are always equal (i.e., the expected utilities of C and D for an agent with degree of guilt aversion equal to one are equal, for every possible probability distribution over the opponent's strategy). Thus, an agent with degree of guilt aversion equal to one will always play in a random way.

Rawls' ideality

Analogously, we now use Rawls' algorithm of ideality the agent's U^*, keeping all the experiment parameters at their default values. Figure 2 shows a schematic representation of the results, organized similarly to previous figure 1.

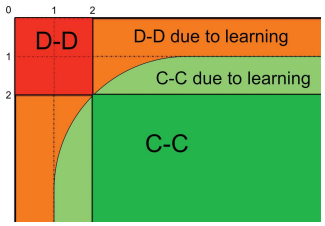

Fig. 2. Rawl's ideality results

Unsimilarly to the results obtained with Harsanyi's ideality, we observe that in this case guilt aversion plays an important role in the emergence of fairness as a moral value. Indeed, agents with a higher degree of risk aversion have a higher average payoff than those who do not present guilt aversion (that is, those who do not experience guilt feelings triggered by the transgression of the fairness value).

In addition, in Rawls' case we have some nuances that we did not have in Harsanyi's case. In particular, the strategy chosen by the agents not only depends on their degree of guilt aversion, but also on the number of interactions they have been part of. This highlights the influence of the learning curve on the decision-making process.

4 Conclusions and Future work

Hence we can finish by concluding that, from an evolutionary point of view, it is the moral values *à la* Rawls that emerge and that guilt aversion does play an important role in the suitability over time of fairness. As R. Frank stated, "[t]he fact that it might sometimes be best to ignore moral emotions does not imply that it is always, or even usually, best to ignore them. If we are to think clearly about the role of moral emotions in moral choice, we must consider the problems that these emotions were molded by natural selection to solve. Most interesting moral questions concern actions the individual would prefer to take except for the possibility of causing undue harm to others. Unbridled pursuit of self-interest often results in worse outcomes for everyone. In such situations, an effective moral system curbs self-interest for the common good. [Furthermore,] moral systems must not only identify which action is right, they must also provide motives for taking that action." [13, pp. 7-8]

This article presents an on-going piece of research that is much to be completed. Although many, we shall note here some of the possible research questions to investigate further. Firstly, we would like to test the model with different population distri-

butions (according to their degree of guilt aversion), in order to explore which is the minimum percentage of guilt averse agents that is necessary for sustaining cooperation. Secondly, in the present model, an agents' learning algorithm is specific for every agent with whom they have interacted; that is, there is no 'global' learning, in the sense of an intuition of the global trend to cooperate or to defect of the rest of the population. Furthermore, it would also be interesting to analyze the results for a case-scenario were the degree of guilt aversion of the agents is perceivable by their opponents (cf. [13]).

5 Acknowledgement

This work is part of the EmoTES ("Emotions in strategic interaction : theory, experiments, logical and computational studies") research project. The EmoTES project is funded by the French National Research Agency (ANR).

References

1. I. Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.
2. R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.
3. P. Battigalli and M. Dufwenberg. Guilt in games. *The American Economic Review*, 97(2):170–176, 2007.
4. R. F. Baumeister, A. M. Stillwell, and T. F. Heatherton. Guilt: An interpersonal approach. *Psychological Bulletin*, 115:243–267, 1994.
5. R. Benabou and J. Tirole. Incentives and prosocial behavior. *The American Economic Review*, 96(5):1652–1678, 2006.
6. K. Binmore. *Natural justice*. Oxford University Press, New York, 2005.
7. S. Bowles and H. Gintis. Prosocial emotions. Working Paper 02-07-028, Santa Fe Institute, 2003.
8. G. W. Brown. Iterative solutions of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York, 1951.
9. G. Charness and M. Dufwenberg. Promises and partnership. *Econometrica*, 74:1579–1601, 2006.
10. Y. Cheung and D. Friedman. A comparison of learning and replicator dynamics using experimental data. *Journal of Economic Behavior and Organization*, 35:263–280, 1998.
11. R. Dawkins. *The God Delusion*. Houghton Mifflin Co., Boston, 2006.
12. E. Fehr and S. Gaechter. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181, 2000.
13. R. H. Frank. The Status of Moral Emotions in Consequentialist Moral Reasoning. In P. J. Zak, editor, *Moral Markets: the critical role of values in the economy*. Princeton University Press, 2007.
14. D. Fudenberg and D. Levine. *Learning in Games*. MIT Press, Cambridge, 1998.
15. J. Halpern and R. Pass. Iterated regret minimization: a new solution concept. *Games and Economic Behavior*, 74(1):194–207, 2012.
16. J. C. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63:309–321, 1955.

17. J. C. Harsanyi. Morality and the theory of rational behaviour. In A. K. Sen and B. Williams, editors, *Utilitarianism and beyond*, pages 39–62. Cambridge University Press, Cambridge, 1982.
18. G. Loomes and R. Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824, 1982.
19. G. Loomes and R. Sugden. Testing for regret and disappointment in choice under uncertainty. *The Economic Journal*, 97:118–129, 1987.
20. J. J. Mansbridge. *Beyond self-interest*. University of Chicago Press, 1990.
21. W. Poundstone. *Prisoner's Dilemma*. Doubleday, NY NY, 1992.
22. J. Rawls. *A theory of Justice*. Harvard University Press, Cambridge, 1971.
23. J. Robinson. An iterative method of solving a game. *Annals of Mathematics*, 54(2):296–301, 1951.
24. J. Searle. *Rationality in Action*. MIT Press, Cambridge, 2001.
25. A. Sen. *Collective choice and social welfare*. Holden-Day, San Francisco, 1970.
26. P. Taillandier, D.-A. Vo, E. Amouroux, and A. Drogoul. GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In *Proceedings of PRIMA'10*, volume 7057 of *LNCS*, pages 242–258. Springer, 2012.
27. J. Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, 1995.