

Author's Accepted Manuscript

Bargaining Models of Depression and Evolution of Cooperation

Tom Rosenström



www.elsevier.com/locate/jtbi

PII: S0022-5193(13)00165-3
DOI: <http://dx.doi.org/10.1016/j.jtbi.2013.04.012>
Reference: YJTBI7274

To appear in: *Journal of Theoretical Biology*

Received date: 1 October 2012
Revised date: 3 April 2013
Accepted date: 5 April 2013

Cite this article as: Tom Rosenström, Bargaining Models of Depression and Evolution of Cooperation, *Journal of Theoretical Biology*, <http://dx.doi.org/10.1016/j.jtbi.2013.04.012>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Bargaining Models of Depression and Evolution of Cooperation

Tom Rosenström

IBS, Unit of Personality, Work and Health Psychology, University of Helsinki, Finland

Keywords: evolution; cooperation; depression; adaptation; dynamical systems; social networks

Corresponding author:

Tom Rosenström: IBS, Unit of Personality, Work and Health Psychology, University of Helsinki (Siltavuorenpenger 1 A), P.O.Box 9, 00014, Helsinki, Finland. Tel: +358 9 1912 9396, Fax: +358 9 1912 9521, Email: tom.rosenstrom@helsinki.fi

Abstract

This paper presents a model for evolutionary origins of unipolar depressive disorders and reviews empirical evidence in support of the “nonparticipation hypothesis of depression” suggested by the model. In line with previous theory, depression is interpreted to provide means for not participating in joint enterprises that would otherwise be difficult to avoid (‘bargaining’ model); therefore, it provides an additional alternative strategy to social-imitation dynamics occurring in Public Good Games. Average income from joint enterprises modeled with a Public Good Game is interpreted to drive group-level genetic selection. Based-on these assumptions, it is shown that if a joint enterprise is sufficiently important for both the between-individual and the between-groups competition, then a nonparticipation strategy (or gene for depression vulnerability) will be present in total population almost surely (i.e., with probability one). The modeling framework serves to explain several other seemingly unrelated empirical observations, such as association between income inequality and depression, co-morbidity and proximal biological mechanisms for negative emotions, and mechanisms for the social network-dynamics of emotion.

1. Introduction

Unipolar depressive disorder is one of the leading global causes of disability-adjusted life-years (Murray and Lopez, 1997; Mathers and Loncar 2006). In fact, depression is so commonly observed across the world that an increasing number of researchers have attempted to explain it from the perspective of evolutionary adaptations (Nesse, 1991, 2000;

Watson and Andrews, 2002; Andrews and Thomson Jr, 2009; Gilbert, 2006; Price et al., 2004; Sloman, 2008). The adaptive nature of severe depression has been questioned, however (Nettle, 2004). It is difficult to see a prolonged complete cessation of action as an adaptive response to any situation; scientific adaptation arguments do not mesh with clinicians' intuition (Friedman, 2012), and they often suggest views that are incompatible with some subsets of the common depressive symptoms (Hagen, 2003, 2011). One of the most consistent evolution-related models is the *Bargaining Model* of depression that is concerned with social interactions among people; it proposes that one of the central functions of depression is to “withhold benefits from others until better terms are forthcoming” (Hagen, 2003). The lack of interest and inactivity that occur in depression deliver a “message” to a social group that for the depression sufferer there is little or no difference in fitness benefits obtained by investing heavily to current joint enterprises. If the depressed individual could provide something for the group, this “something” is withheld until better terms of cooperation are negotiated. Such a bargaining strategy can be effective when more direct coercion or switching of social groups are not.

The Bargaining Model can explain symptoms that are difficult for other evolutionary accounts. For example, the analytical rumination hypothesis posits that depression is an adaptation that promotes cautious analysis of difficult, fitness-related problems (Andrews and Thomson Jr, 2009), but it is difficult to explain suicidal ideations from the viewpoint of cognitive re-evaluation of ones goals. From the viewpoint of the Bargaining Model, a completed suicide would most clearly remove an individual as a source of valuable benefits to others; suicide attempts are necessary to underwrite the creditability of the threats, and the completed ones are a cost to maintain creditability (Hagen, 2003). An evolution of suicide signaling/bargaining strategy is feasible when the rate of threats is much higher than the rate of attempts, and the latter is still much higher than the rate of completion;

the average benefit received from influencing others over many generations by genes encoding for this strategy could exceed the cost suffered by those genes from the deaths by suicide (Hagen, 2003). Some adaptive theories have interpreted depression as an appeasement display designed to de-escalate a social conflict situation (Price et al., 2004); yet, what appeasement is needed, say, in a case where depression derives from having a difficult child or a dying spouse? Instead, a difficult child or a dying spouse can be taxing for a caregiver, and depression may serve to bargain the help of the wider social group (Hagen, 1999, 2003). More examples can be found elsewhere (Hagen, 1999, 2003, 2011; Watson and Andrews, 2002).

Although much has been written about depression as an individual-level adaptation, the present author is not aware of previous studies that examine the meaning of depression for the adaptive fitness of a group of individuals, or society; even though depressive symptoms have been shown to spread in social networks (Rosenquist et al., 2010, Hill et al., 2010). The present report outlines the implications of Bargaining Model for the emergence of cooperation under social-imitation dynamics (Nowak, 2006, 2012; Sigmund, 2010) and for the resulting fitness of groups of individuals. It is argued that by combining the Bargaining Model with models for evolution of cooperation, a more comprehensive theory for adaptive origins of depression is achieved.

An option to abstain from a joint enterprise has been shown to be critical for the sustained cooperation in joint efforts of many self-interested agents (Fowler, 2005; Hauert et al., 2007; De Silva et al., 2010; Sigmund et al., 2010). The Bargaining Model (Hagen, 2003) and similar accounts suggest that depression offers a way to not participate in a joint enterprise while, at the same time, retaining critical baseline benefits. Therefore, the combined hypothesis (detailed in the next section) is called the *nonparticipation hypothesis of depression*, in short, the nonparticipation hypothesis. The considerable within-individual

time-variation (Hyde, et al., 2008; Rosenström et al., 2013) in depression will be explained by social imitation dynamics (Sigmund, 2010) and the genetic sensitivity for environment (Caspi et al, 2010; Karg et al., 2011) will be explained by group-level genetic selection (Nowak, 2006, 2012).

The main contribution of this study is the provision of a new hypothesis for the adaptive origins of depression that incorporates genetic, environmental, and dynamic social effects. Therefore, it contributes to the complex task of determining “which depressions are adaptive in which context” (Gilbert, 2006). Importantly, it is not implied that depression would necessarily be adaptive for a single individual suffering from it, or for a single loved one of a depressed person. The author merely suggests that, because of the naturally occurring social dynamics, it is adaptive for a group of individuals that some or all have a capacity for depression. The nonparticipation hypothesis is not in direct conflict with existing individual-based theories of depression, but can co-exist with them. Clinically defined ‘major depressive disorder’ is not differentiated from low moods and depressive symptoms here, unless mentioned. Depression appears to be a continuum (Hankin et al., 2005; Solomon et al., 2001, Jokela et al. 2011), where discrete diagnosis is typically justified by a lack of understandable cause or function for the symptoms or by their severity (Hagen, 2011); the present discussion proceeds on the level of abstraction where precise clinical definitions add little value.

The next section describes how existing models for the evolution of cooperation lead to the nonparticipation hypothesis of depression, and sets up a formal model. Then, the results section shows how this model implies that group-level selection (Nowak, 2012) favors the emergence of depression-sensitivity genes. In the discussion section, empirical research in support of the hypothesis is reviewed, and the report is concluded. Altogether, the non-participation hypothesis yields functional explanations for a diverse set of seemingly

unrelated empirical observations, including co-morbidity among negative emotions, cascading of emotions in social networks, and connection between income inequality and depression. Alternatives for the previous “mainstream approach” to depression (Hagen, 2011) are now sought for, as it has fallen under heavy critique (Hagen, 2011; Andrews, 2012; Kleinman, 2012a), and standard somatic interventions may incur increased rates of mortality, suicide attempts, strokes, epilepsy, falls, fractures, and other adversities in the elderly (Coupland et al., 2011).

2. The model and its rationale

Evolutionary adaptations have been defined as “inherited and reliably developing characteristics that came into existence through natural selection because they helped to solve problems of survival or reproduction during the period of their evolution” (Buss, 2008; Buss et al., 1998). The underlying assumption of this study is that, during history, cooperation among bands of humans has solved problems of survival. This is a notion that many concur with, starting from Darwin: “There can be no doubt that tribe including many members who ... were always ready to give aid to each other and sacrifice themselves for the common good, would be victorious over other tribes; and this would be natural selection” (Darwin, 1871; Nowak, 2012; Sigmund, 2010, p. 151). Yet, natural selection cares not of the intent, but of the outcome; cooperation arising from immediate selfish interests can be as efficient adaptation as cooperation due to other intentions, and may be easier to establish among strangers than self-sacrifice. Striving for the simplest dynamic model to begin with, this study is concerned with selfish motivations as a driving force of social dynamics and depression [e.g., as opposed to kin-selection (Nowak, 2012; Sigmund, 2010)].

Many people seem willing to contribute effort for the punishment of others in order to maintain the cooperation of majority (e.g., maintain jailhouses). In joint enterprises, some should invest their effort to render free-riding and exploitative actions unprofitable for the others. A social dilemma is why to make an effort to punish when profit from the joint enterprise will be equally distributed among participants anyway? A punisher needs to invest more than a by-stander who equally benefits from the punisher's efforts; punishing does not seem to directly increase any individuals fitness compared to others. Therefore, the rationale behind the emergence of 'altruistic' costly punishment is an evolutionary puzzle intimately linked with the emergence of cooperation (Fowler, 2005; Coleman, 2006; Axelrod and Hamilton, 1981). A simple ingredient allowing the emergence of costly punishment in a joint enterprise is an option to abstain from the enterprise (Fowler, 2005; Hauert et al., 2007; De Silva et al., 2010; Sigmund et al., 2010). In this section, a model of joint enterprises by Hauert and others (2007) is first shortly reviewed, and then a connection between depression and abstaining from a joint enterprise is established. Finally, the emergence of a 'depression gene' is modeled as a function of the incremental value of the nonparticipation/depression strategy for the group's gain from joint enterprises.

2.1 A model for joint enterprise

When attempting to explain the emergence of cooperative behavior in joint enterprises, it is customary to begin with agents/individuals who attempt to optimize some selfish gain, and then derive the conditions where the selfish effort leads to cooperation among the agents (Sigmund, 2010; Nowak 2006; Axelrod and Hamilton, 1981). A Public Good Game consists of several rounds of interaction among subsets of individuals from a population of M individuals; a well-behaved infinite-population equivalent is not available for the particular

four-strategy game we are interested in (Hauert et al., 2007, Sigmund, 2010). In each round, N players are randomly drawn from the population of M individuals and given an option to participate in a joint enterprise, where they can choose to invest a cost/effort c . The efforts are then multiplied by the factor r ($r > 1$), yielding the benefit rc that is evenly distributed among those participating in the enterprise—regardless of who paid the cost. After each round, there is an option to pay more in order to impose a punishment for those who participated, reaped the benefits, but did not contribute the payment c for the public good.

The population of M individuals consists of players applying one of the four alternative strategies: M_c Cooperators pay the cost in each round; M_d Defectors (the free-riders) never pay; M_p Punishers pay the cost in each round, and also pay γ per each Defector in order to impose an unavoidable fine of βM_p for the Defectors (where $\beta > \gamma$); finally, M_n Nonparticipants do not take part into the game, but gain an amount σ per round independently of the game (if $N-1$ Non-participants happen to be selected, the enterprise is cancelled from that round, and the remaining participant also gains σ). The four strategies are the only alternatives: $M_c + M_d + M_p + M_n = M$. For the game to be both potentially lucrative and such that nonparticipation is a real option, inequality $0 < \sigma < (r - 1)c$ must hold; that is, when all players cooperate it is advantageous to participate, when all defect a Nonparticipant does better. For non-triviality, the game cannot be beneficial regardless of other players actions (i.e., $cr/N - c < \sigma$ must hold).

Models for Public Good Games usually apply fitness-based *imitation dynamics*: successful strategies are imitated by other players in the population with intensity proportional to the product of relative fitness of strategy and the frequency of the players using that strategy (Sigmund, 2010). The frequency of a given strategy in the population can be interpreted as the frequency for observing the associated behavioral model, which is directly related to occasions where “imitation” of that strategy can occur. Because the fitness

of a strategy usually depends on the distribution of strategies among the players, fitnesses of strategies change with the proportion of participants using them. Therefore, imitation can give rise to rich time-dynamics. In both evolutionary and imitation dynamics, it is typically assumed that rare ‘innovations’ exist; that is, a randomly chosen strategy is introduced in rare occasions via behavioral exploration or mutation (Nowak, 2006; Sigmund, 2010). A small rate of innovations is assumed here too.

Some examples of how proportions of played strategies affect their fitness are apparent: If most co-players are Defectors, it is wisest to be a Nonparticipant and gain σ . A lone Punisher would need to punish left and right, imposing huge costs for the self. A Defector would gain nothing, and a Cooperator would lose $c(1 - r/N)$. Therefore, imitation dynamics will favor Nonparticipants. But as they take over the population, the game becomes most lucrative for Cooperators. Again, when Cooperators are a majority, the incentive to defect is obvious: the gain is proportional to rc/N compared to the Cooperators’ $rc/N - c$. This type of oscillation of three strategies is known as ‘Rock-Paper-Scissors’ dynamics (Sigmund, 2010; Nowak, 2006); no one benefits more than the Non-participants’ σ in the long run [the time-average of the pay-offs is σ (Sigmund, 2010, p. 136)]. When participation in the game is compulsory; only Cooperators, Defectors, and (costly) Punishers play; Defectors dominate the time-average of the strategy-distribution (Hauert et al., 2007; De Silva et al., 2010; Sigmund, 2010); the average pay-off in the population is less than σ . Therefore, the game cannot be simplified without compromising the adaptive value of the joint enterprise for the population. Only when all the four strategies are present, the collaborative effort flourishes most of the time (Hauert et al., 2007; De Silva et al., 2010; Sigmund, 2010).

In principle, one could conceive many more possible strategies and game-settings; however, the above model is simple and its predictions show remarkable robustness for alternative formulations (Hauert et al., 2007; De Silva et al., 2010; Sigmund et al., 2010;

Sigmund, 2010). For example, punishing can be implemented via “sanctioning institution” supported by the Punishers instead of the above-defined “peer-punishing” without changing the essential result regarding the importance of the voluntary participation (Sigmund et al., 2010). The option to abstain from the joint enterprise is necessary for the coerced cooperation to emerge in most model formulations. For this reason, and because it is so difficult to find simple dynamical settings that encourage cooperation over defection, the discussed four-strategy model is most likely to have wide-ranging biological and social relevance. The next section outlines how depression could be interpreted as nonparticipation, and why this might have been crucial during the evolution of the cooperative human.

2.2 Non-participation via depression

Throughout the evolutionary past, humans have been highly dependent on cooperative efforts (Buss, 2008). For example, in small bands of hunter-gatherers it has been adaptive to share the uncertainty in food acquiring. The temptation to defect has obviously existed in the past, and may have even resulted in genetic adaptations (Buss, 2008). The cooperative and punishing behaviors have certainly existed in the past. However, in close-knit communities, the option to not participate may sometimes have been less evident. Assuming that a lone human faces difficulties in surviving, the option to not ‘play’ cannot be taken as granted (MacDonald and Leary, 2005). In above section, however, it was discussed that the option to abstain is crucial for the emergence of constructive cooperation.

Depression may have offered a means to abstain from a joint enterprise while simultaneously avoiding enforcement from others. Depressed persons suffer from fatigue, inefficiency, and lack of effort. Even brief interactions with a depressed person suffice to convince others that they are not going to get the kind of relationship they want (Coyne,

1976; Boswell and Murray, 1981). Normally this would run a risk of punishment or ostracism by the group, but the depressed compensate by eliciting empathy in others: “The symptoms of depressed person are aversive yet powerful in the ability to arouse guilt in others and to inhibit any direct expression of annoyance and hostility from others” (Coyne, 1976) and people are less likely to retaliate against depressed than cheerful person (Surbey and Simpson, 2010). Depression may thus be an effective way to not participate, and still maintain modest benefits (the Nonparticipants’ σ). Choosing the modest Nonparticipants’ share is bargaining in the precise sense of the Bargaining Model; that is, one withdraws his share of effort from the ‘Defectors’ until more cooperative terms of play emerge (Hagen, 2003). Even though bargaining were not a conscious act and no direct benefit would come for the depressed person nor any other single individual, depression as a phenomenon may have benefitted groups of individuals engaging in joint enterprises by promoting the emergence of cooperation in social dynamics (Hauert et al., 2007; De Silva et al., 2010; Sigmund, 2010; Sigmund et al., 2010).

Depressed mood can instantiate a nonparticipation strategy more effectively than many other adverse conditions. For example, schizophrenia also elicits aversive reactions in healthy people (Boswell and Murray, 1981), but a schizophrenic does not readily switch strategies and start fully re-participating in a joint enterprise; it is not as dynamic condition as depression (Bienvenu et al., 2011). Depression and low moods preserve potential for recovery/change, as required in imitation dynamics. In addition to withdrawal, depressed mood is thought to signal submission in a conflict situation; that is, it carries a behavioral message “I will not retaliate your actions” (Price et al., 2004). Therefore, the depressed are not easily mistaken by others as applying a punishing strategy, despite their obvious dissatisfaction regarding a state of affairs. Instead, schizophrenia (causing, e.g., blunted affect) does not appear designed for eliciting the empathy of others, and therefore it lacks an

obvious mechanism that would guarantee any Nonparticipants' pay-off. Involved delusions and disorganized behavior may rather cause social conflicts than aid in avoiding them.

In existing simulations of the four-strategy Public Good Game, the population is on average in a state where the proportion of Nonparticipants in the population ranges approximately from 0% up to 20%, depending on other parameters (Hauert et al., 2007; De Silva et al., 2010; Sigmund et al., 2010). This is roughly in the same ball-park with epidemiologic prevalence estimates for the major depressive disorder that may vary from 1.5% in Taiwan to 19.0% in Beirut, Lebanon (Weissman et al., 1996; Kessler et al., 2003); more precise prevalence estimates provide little theoretical value as the estimates depend on diagnostic criteria that are under active critique and reconsideration (Hagen, 2011; Kleinman, 2012b). Since depression could feasibly serve as means for nonparticipation, the next section describes a simple model aiming to answer whether or not natural selection on a group-level leads to a gene/-s for depression vulnerability (Caspi et al., 2010; Bienvenu et al., 2011; Duncan and Keller, 2011; Karg et al., 2011).

2.3 Group-level selection driven by joint enterprise outcomes: the model

The approximate correspondence between the estimates for depression and 'nonparticipation' prevalence that was mentioned above was derived for a closed system of four strategies. The scientific premise of parsimony requires a justification for each and every added layer of complexity in a model. Taking cooperative, defective, and punishing strategies as given, this study aims to demonstrate that evolution will favor the emergence of the fourth behavioral strategy of nonparticipation; behavioral strategy refers here to a robust and general mode of behavior that has a biological basis, not only to an abstract conscious strategy. Conscious cortical processing does not provide very efficient means for coping with important biosocial

threats, as can be seen, for example, in the frequent injuries suffered by those who cannot sense pain or discomfort (Nesse, 1991). Furthermore, evolution must build on existing ‘facilities’. Indeed, it is often difficult or impossible to separate emotional and physical pain (Lee, et al., 2007; MacDonald and Leary, 2005), and for example, it has been argued that this is because the physical pain system “was already in place when social animals evolved adaptations for responding social exclusion” (MacDonald and Leary, 2005).

According to present model, gains from the above described joint enterprises drive a natural selection process between depression-vulnerability gene (nonparticipation strategy) existing in population *versus* not existing; the idea is that such a gene is introduced to populations on rare occasions (with a low mutation rate) and affects a competition against other populations, and therefore the survival of the population. Individuals in each group/population occasionally engage to a joint enterprise, the time-average of the groups average income from that enterprise affects the group’s genetic fitness proportionally to the selection strength, s_g . Selection strength is a parameter that describes how important the enterprise is for natural selection; how much it biases natural genetic drift in a Moran process towards one or other outcome (Nowak, 2006; Sigmund, 2010). The modeled ‘gene’ is for availability of, or potential for, the fourth nonparticipation strategy, whose final expression in population depends on the ‘environment’ of other strategies. Therefore, the modeled totality represents an interaction of genes and environment.

Hauert and others (Hauert et al., 2007; Sigmund, 2010) computed how the above described joint enterprise progresses on finite groups under imitation dynamics and low rate of innovations (i.e., at a low rate, someone spontaneously tries some other alternative than the prevailing strategy); in such a setting, the time-scales between imitation dynamics and innovations can be separated, which greatly simplifies the analysis. The group always drifts to a pure state where only one strategy is in use, and when an innovation randomly

introduces one of the other three strategies, it competes with the dominant one according to a two-strategy Moran process. Because the competition resolves much faster than the next innovation occurs, the innovation- and competition time-scales can be separated and the resulting process is a four state Markov Chain whose stationary distribution is given by certain eigenvector of a transition matrix formed from the fixation probabilities of the two-strategy pairings (Hauert et al., 2007; Sigmund, 2010). The stationary distribution yields the time-average of each strategy in a group with potential for all four strategies (Cooperators, Defectors, Nonparticipants, and Punishers). Similar time-averages can be also computed for groups with potential for only three strategies: Cooperators, Defectors, and Punishers. These latter kinds of groups can be interpreted to lack a Nonparticipation strategy, or a depression gene/-s.

Supposing that competition between groups with and without a depression gene exists in a larger population of groups, then the fixation probability of the depression gene is defined as the probability that a single group with depression gene gradually takes over the whole population of M_g groups, where M_g-1 groups do not originally have that gene; in the end, all groups have been replaced by an equal amount of groups with all having the gene. This reflects the situation of where a new ‘mutation’ spreads to a population, mutation in genetic dynamics being analogous to ‘innovation’ in imitation dynamics. It is customary to consider fixed amount of groups and a Moran process for their evolving genetic status (Nowak, 2006; Sigmund, 2010). In a (general) Moran process, one group is chosen for extinction and one for replication in each round, with the choice for replication versus extinction being based on the ratios between functions of the two groups’ genetic fitnesses, F and G ; more precisely, on the ratio between $1-s_g+s_gF$ and $1-s_g+s_gG$, where s_g ($0 \leq s_g \leq 1$) is the selection strength of environment for the gene in question (Nowak, 2006; Sigmund, 2010).

Again, when mutations (cf. innovations) are rare, the Moran process becomes a two-state Markov Chain with transition probabilities equaling the fixation probabilities,

$$\rho_{1,2} = 1 / \left(1 + \sum_{k=1}^{M_g-1} \prod_{i=1}^k \frac{1 - s_g + s_g G}{1 - s_g + s_g F} \right),$$

where $\rho_{1,2}$ is transition probability from ‘no-gene’ state to a state with the ‘gene’, $1 - \rho_{1,2}$ is the probability of staying in the *no-gene* state (mutation introduces the gene, but it goes extinct before taking over the population), $\rho_{2,1}$ is transition probability from *gene* state to *no-gene* state (a mutation deletes the gene from population), and $1 - \rho_{2,1}$ is the probability of staying in *gene* state (deletion does not take over). Again the eigenvector of the standard Markov-Chain theory transition matrix that is associated with the maximal eigenvalue yields the average times that the chain spends in each state, or probabilities of finding the chain at a given state in a randomly chosen time: $P(\text{gene}) = \rho_{1,2}/(\rho_{1,2} + \rho_{2,1})$, and $P(\text{no gene}) = \rho_{2,1}/(\rho_{1,2} + \rho_{2,1})$.

The model assumes that imitation dynamics of the Public Good Game of Hauert and others (2007) applies in the within-group level, and that the genetic selection between groups with and without the depression gene occurs on much slower time scales. Imitation process *versus* innovation occurrence time-scales can be separated for the faster within-group dynamics, and natural-selection process *versus* mutation-occurrence time scales separate for the slower progressing genetic dynamics. For low innovation/mutation rates, such “adiabatic” approximations closely correspond to simulation results (Hauert, 2007; Sigmund, 2010).

When $(\pi_c, \pi_d, \pi_n, \pi_p)$ is the stationary distribution (relative time-averages) of cooperative, defective, nonparticipating, and punishing strategies in the Public Good Game with the depression gene, and $c = 1$, the time-averaged outcome from that join enterprise is:

$$F^* = \pi_c(r-1) + \pi_d 0 + \pi_n \sigma + \pi_p(r-1) = (\pi_c + \pi_p)(r-1) + \pi_n \sigma.$$

When nonparticipating strategy is not available and (π'_c, π'_d, π'_p) is the stationary distribution of the three available strategies in such joint enterprise, the average pay-off is:

$$G^* = (\pi'_c + \pi'_p)(r-1).$$

The stationary distributions $(\pi_c, \pi_d, \pi_n, \pi_p)$ and (π'_c, π'_d, π'_p) are derived exactly as in previous reports (Hauert, 2007; Sigmund, 2010), and re-stating the formulas are therefore avoided. We normalize as, $F = F^*/(F^* + G^*)$ and $G = G^*/(F^* + G^*)$, so that s_g can be directly interpreted as the relative selection strength of environment for the depression gene (i.e., for availability of a nonparticipation strategy). The success of strategies compared to other individuals' strategies within a group affects the intensity of imitation proportionally to the imitation (strategy-selection) strength denoted by s_i [corresponds to the selection strength parameter s in previous studies of the Public Good Game (Hauert et al., 2007; Sigmund, 2010)].

The primary outcome of interest in this study is the probability (relative time-average) with which a depression gene, or a nonparticipating strategy, is found from the wider population (among all groups) compared to the state without such gene; that is, the quantity $\rho_{1,2}/(\rho_{1,2} + \rho_{2,1})$. In addition to calculations that straightforwardly extend previous reports (Results section), further empirical evidence in support of the results and the modeling framework will be reviewed (Discussion section). All computations and figure drawing was performed using Matlab[®] software (Mathworks, Natick, Massachusetts, USA) version 7.10.0.499 (R2010a).

3. Results

Figure 1 shows the average system state (stationary distributions of strategies) in the Public Good Game as a function of imitation-selection strength s_i ; the lower-right panel also shows average productivity of the joint enterprise in a group with all four strategies available, in a group that lacks the Punisher strategy, and in a group that lacks the Nonparticipants.

Parameters were set as in previous work, and obtaining the exact same results served to ensure that the present model-implementation concurs with the previous ones (Hauert et al., 2007; Sigmund, 2010). When the joint enterprise was irrelevant for strategy imitation ($s_i = 0$), all strategies were equally frequently observed in all models. When s_i grew, some strategies were preferred by individuals within a group more often than other strategies.

When all four strategies (Cooperators, Defectors, Nonparticipants, Punishers) were available, the conditionally cooperative Punishing strategy was found most often within a group. When punishing strategy was lacking, rock-paper-scissor dynamics settled in, and the three remaining strategies oscillate with more even proportions. When the nonparticipating strategy was lacking, Defectors took over the population for moderate to large s_i . For a very small s_i , when joint enterprise was not very significant for individual players, income of the joint enterprise was largest for three-strategy game including Cooperators, Defectors, and Punishers. When s_i grew, however, the importance of the option to abstain was increasingly reflected in the average income, and therefore in the fitness of groups (in the parameters F and G).

Analyzing the setting without Punishers, where the infinite group-size analysis is not unstable, provides some insight to the role of Nonparticipants (Sigmund, 2010); namely, nonparticipation is a response to defection. When ‘unfair’ sharing of spoils of the joint enterprise increases in a group, this is followed by an increasing prevalence of nonparticipation (see Figure 2). Many Nonparticipants make cooperation a comparably successful strategy, until defecting strategies arise again: the game is locked into a fruitless

rock-paper-scissors dynamics, where the time-average of total income equals the Nonparticipant's σ (Sigmund, 2010). The figure is based on equations derived by Sigmund (2010), and is provided for purposes of discussion section; because of the intuition it may provide for the reader.

Figure 3 shows the probability of all groups (entire population) having a Nonparticipation strategy (a depression-sensitivity gene) in their pool of available strategies; that is, the figure shows the average time the Moran process spends in a state where depression gene exists as opposed to state where it does not exist, provided that the incomes of a group's joint enterprises are reflected upon that group's fitness. The probability of depression gene(s) is plotted as a function of the importance of the joint-enterprise outcome for individuals (imitation-selection strength s_i in social dynamics) and importance of outcome for the group's survival (evolutionary selection strength s_g among groups/species/tribes/etc.). The results show that, for most selection-strength values, the probability of depression gene(s) grows as a function of both the number of competing groups and the number individuals per group; with sufficiently many groups and sufficiently many individuals in them, depression gene is a certainty imposed by natural selection under the modeled dynamics. When the importance of joint enterprises is very low for involved individuals ($s_i \leq 0.006$) but high for the groups, the pay-off can be higher for groups without the Nonparticipants (Figure 1), implying that selection favors groups without a depression gene(s) (Figure 3).

4. Discussion

This study showed that when outcomes of joint enterprises drive a social competition among individuals within groups *and* an evolutionary competition among a finite number of groups,

evolution is likely to introduce some means for individuals to avoid participating joint enterprises. More precisely, if the joint enterprise is sufficiently important for both the between-individuals and the between-groups competition, then a Nonparticipation strategy will be present almost surely (with probability one). The present work interpreted depression as a strategy that has provided a way to not participate to otherwise nearly compulsory joint enterprises during evolutionary history. The analysis is not intended to belittle the individual-level suffering among the depressed, but to provide a sought for addition to the scientific understanding of evolutionary origins and group-dynamical aspects of depression.

Several theories of depression have interpreted it as means for avoiding participation to joint enterprises with undesirable terms (e.g., Hagen, 1999, 2003; Watson and Andrews, 2002; Andrews and Thomson Jr, 2009; Price et al., 2004), and some evidence for a genetic sensitivity for depressogenic environments does exist (Caspi et al, 2010; Bienvenu et al., 2011; Karg et al., 2011). Genes that predispose for depression may have helped to ensure the Nonparticipation strategy in join enterprises among groups of people during the evolutionary history. The results obtained in the present study supported this view, but they naturally depend on the modeling assumptions that were made. The following text reviews some additional empirical evidence concurrent with modeling of depression as a Nonparticipation strategy in a Public Good Game (i.e., in joint enterprises).

To begin with, an analysis without the Punishers clearly showed that Nonparticipation is a response to defection. Therefore, when 'unfair' sharing of spoils of a joint enterprise increases in a population, this is followed by an increasing prevalence of Nonparticipation (Figure 2). If the situation is such that an alternative enterprise with comparable gains is not readily available, or players are forced to participate, then depression is hypothesized to emerge as a substitute for nonparticipation. In general, a positive correlation

between unfair social situations and depression is expected. Such empirical correlations can be found.

Income inequalities offer one example where ‘unfair’ sharing may be involved, and clear Punishing strategy may not be available: everyone is required to participate in the joint enterprise of economy (other ‘nonparticipation’ strategies are not readily available), high income inequality implies that some (‘Defectors’) benefit from the work of others (‘Cooperators’), and making profit is usually legally accepted, even encouraged (no ‘Punishers’). Therefore, imitation dynamics predicts that disproportionate increases in income inequality are followed by increases in depression. Indeed, income inequality is more clearly associated with mental health than with the physical aspects of health (Muramatsu, 2003).

Village-level income inequality among foraging-farming societies in the Bolivian Amazon (Tsimane’) is associated with more negative emotions, independently of social capital, residential segregation, and public policies (Godoy et al., 2006). Many other studies also report associations between income inequality and depression or mental health (Ahern and Galea, 2006; Fiscella and Franks, 2000; Weich et al., 2001). Depression increases work absenteeism and early retirement, and decreases productivity (Karpansalo et al., 2005; Benden and Farvolden, 2010). Therefore, ‘unfair’ conditions (presence of defection) seem to promote depression, and the depressed do not fully participate in the joint enterprises of the society. Generally, unequal returns for equal efforts lead to avoidance of the joint enterprise in question; even monkeys refuse to participate if they witness that other monkeys obtain a more attractive reward for equal effort (Brosnan and De Waal, 2003).

A study from Britain found that income inequality was more strongly associated with worse mental health in those with higher absolute income levels (Weich et al., 2001),

which may explain null findings in a study that oversampled poor individuals (Sturm and Gresenz, 2002). As countries get richer, rates of mental illness increase; prevalence of mental illness being correlated with income inequality (Pickett et al., 2006). In the current modeling context this concurs with a following observation: the lower the Nonparticipants' pay-off, the more tempting the other strategies (Sigmund, 2010). As an exception to the above observations, lower income groups had clearer association between neighborhood-level income inequality and depression in a post-disaster context (Ahern and Galea, 2006).

Although it seems more likely that inequalities cause depression than the other way around, evaluation of a dynamic model would naturally benefit from accurate time-series. In the United States of America, economic inequality grew during the early 1980s, stagnated in the late 1980s, and re-accelerated in the 1990s (Bernstein and Mishel, 1997). Between the years 1987 and 1997, the rate of outpatient treatment for depression increased from 0.73% to 2.33% (Olfson et al., 2002). Thus, at least these numbers are in the temporal order predicted by the nonparticipation hypothesis. Determination of the correct time-scale and reference population, however, may represent a difficulty in the accurate evaluation of the nonparticipation hypothesis of depression. For example, whereas the total global income inequality has decreased, global within-country inequality has grown through 1990-2008 (Clark, 2011), as has the global amount of disability-adjusted years of life due to unipolar depressive disorders (Mathers and Loncar, 2006); it was precisely the within-group situation that was relevant for the social imitation dynamics in the present model.

In addition to income data, work-life in general abounds with joint enterprises. Work conditions where people experience that they gain little rewards (money, esteem, promotion prospects, and job security) compared to efforts they have invested lead to lower mental health later on [depression, anxiety, chronic fatigue, and psychotropic drug consumption (Godin et al., 2005)]. In general, the social 'fair-play' characteristics of a work-

unit appear to be more relevant for depression than work-content characteristics like job demands and decision latitude (Ylipaavalniemi et al., 2005; Netterstrøm et al., 2008).

According to the nonparticipation hypothesis, a significant part of the variance in depressive symptoms should derive from social imitation dynamics. Twin studies find depression to be 31%-42% heritable (Sullivan et al., 2000; Bienvenu et al., 2011), leaving room for other time-accumulating sources of variance (Rosenström et al., 2013). Is there evidence that people imitate behavioral strategies of other people, and does this imitation generalize to long-term emotional states?

In experimental Public Good Games, cooperative and defecting behavior is imitated by others over several rounds of the game with different and unrelated participants than the origin of the behavior (Fowler and Christakis, 2010). Therefore, imitation dynamics exist in experimental Public Good Games. Although large longitudinal observational data sets that contain information regarding social connections and depression are rare, the data from Framingham Heart Study indicates that both happy and depressed mood spreads from one participant to another through their social connections (Rosenquist et al., 2010; Fowler and Christakis, 2008). Confounders and homophily (tendency to seek like-minded company) were controlled for, and the results indicated that depressed mood is causally inducted from one person to another (Rosenquist et al., 2010). Indeed, happy and depressed emotions in the Framingham data can be modeled using models for infectious diseases (Hill et al., 2010).

The Framingham studies of social dynamics leave open the mechanism of ‘infection’, or induction (Hill et al., 2010). They also give rise to a puzzle: why (after controlling for homophily and confounders) significantly less mood induction is observed between spouses than between friends over the course of time (Rosenquist et al., 2010; Fowler and Christakis, 2008)? In the context of the nonparticipation hypothesis, these

findings can be explained: one of the mechanisms of induction would be that the participants adopt their close ones interpretations' regarding the reward value of relevant joint enterprises. This can happen, for example, via direct observation of mood or via conversations. The amount of time spent with a spouse, or closer emotional connection to the spouse in comparison to friends, need not be directly relevant for many of the important joint enterprises that one is engaged with.

The present model predicted that the prevalence of cooperative strategies and the total long-term pay-off of the joint enterprise are higher when all the four strategies are present. Experimental evidence from the context of Public Good Games shows that a possibility to pay in order to punish co-players increases the pay-off in the long-run (Fehr and Gächter, 2000; Gächter et al., 2008). Furthermore, altruistic punishment occurs in various different human populations in different continents, nations, environments, and economies (Henrich et al., 2006). These experiments do not explicitly assess the option to abstain, but surely none are forced to participate in them, as it would be against the ethical standards in human experiments. Presently, it is difficult to directly assess the role depression-related nonparticipation in naturally occurring joint enterprises, but some relevant studies do exist. Depression results in absenteeism from a joint enterprise (Karpansalo et al. 2005; Benden and Farvolden, 2010), and it is negatively associated with interpersonal trust (Lester and Gatto, 1990). Depression is also associated with self-reported lack of self-directedness, high harm avoidance, and low cooperativeness (Farmer et al., 2003; Hansenne et al., 1999; Jylhä and Isometsä, 2006; Marijnissen et al., 2002). These behavioral tendencies serve to promote abstaining from collaborative efforts, and depressed individuals also have less links in social networks than the non-depressed (Rosenquist, et al., 2010). In a direct empirical test, more depressed individuals had fewer intentions to cooperate than less depressed individuals in a

game that replicated the classical Prisoner's Dilemma pay-off structure in socially salient contexts (Surbey, 2011; Surbey and McNally, 1997).

So far we have discussed about evidence for the role of depression in social dynamics, and that this role should imply the existence of a biological evolutionary basis; does direct evidence exist regarding a connection between physiological processes and social processes in the context of joint enterprises? The genetic basis of neurotransmitter serotonin has been claimed to interact with external environment in the etiology of depression (Brown and Harris 2008; Caspi et al., 2010; Keltikangas-Järvinen and Jokela, 2010). However, direct pharmacological manipulation of serotonin levels does not seem to provide very efficient nor immediate general intervention to depression, save maybe the acute and severe cases (Kirsch et al., 2008; Fournier et al., 2010; Hagen, 2011; Vöhringer and Ghaemi, 2011; Andrews et al., 2012). Instead, recent experimental evidence indicates that serotonin modulates social decision-making and the associated brain areas (Crockett, 2009).

In the two-player Ultimatum Game (Sigmund, 2010), an experimenter assigns a certain monetary sum and one player (Proposer) can offer a share of it for the other player (Responder). The Responder can then either accept the proposal, and the sum is split accordingly between the two players, or decline and let the experimenter withdraw the sum from both players. Economical rationality would imply that any share greater than zero should be accepted, as it is better than nothing, but in real experiments subjects frequently decline (Sigmund, 2010). In the Ultimatum Game, experimental subjects with temporarily lowered serotonin levels show an increased rate of rejection for unfair monetary proposals (Crockett, 2009; Crockett et al., 2008). This could be interpreted as an increased willingness to impose an altruistic punishment, or as imposing a message of non-participation. In contrast, pharmacologically induced high serotonin levels make subjects less likely to reject unfair offers by boosting their natural aversion for harming other persons (Crockett et al.,

2010). This may promote an unconditionally-cooperative strategy. A genetically modulated tendency toward the defecting behavior may also exist (Buss, 2008).

The genetics of serotonin has been associated with depression and anxiety on one hand (Brown and Harris, 2008; Caspi et al., 2010; Keltikangas-Järvinen and Jokela, 2010), and with anger and hostility on the other hand (Keltikangas-Järvinen and Jokela, 2010; Carver and Miller, 2006; Merjonen et al., 2011). Regarding evolution, depression has been associated with a de-escalating strategy in competition or conflict between individuals, and anger with an escalating/assertive strategy (Price et al., 2004). In the framework of Public Good Games, depression naturally aligns with a nonparticipating strategy, whereas anger promotes a punishing strategy. Hence, serotonin appears to modulate sensitivity to perceive social unfairness and act upon it, leaving open the choice between nonparticipating strategy and a punishing one. The choice may depend on many other environmental, biological, and cognitive contingencies.

Both temporary pharmacological and temporary psychological induction of negative emotions led to increased rejection of unfair offers in the Ultimatum Game (Crockett et al., 2008; Crockett et al., 2010; Harlé and Sanfey, 2007). However, non-medicated, baseline-depressed participants *accept* more unfair offers than non-depressed participants, especially when high in cardiac vagal control (associated with a stronger ability to self-regulate emotion) (Harlé et al., 2010). This was originally interpreted as prosocial behavior, but another interpretation is that the depressed avoided retaliating and conflict, striving for minimal participation. In the Ultimatum Game, they finished better-off, because they accepted more offers (Harlé et al., 2010). Accepting most offers can function as a way to send a message that “I am not actively participating, nor retaliating”, and in the Ultimatum Game, it certainly leads to the best pay-off because any accepted offer yields more than a rejected one.

The nonparticipation hypothesis explains the high co-morbidity between negative emotions (e.g., depression, anxiety, hostility, and anger): they often co-occur and depend on proximal brain structures, because they are adaptations for solving the same problem—promotion of cooperation in individuals with high self-interest. Had anger evolved without the potential for depression, it might have led to populations with Cooperators, Defectors, and Punishers; shown to result in a dominance of defection (Hauert et al., 2007; De Silva et al., 2010; Sigmund, 2010). Had depression evolved without anger, populations with only Nonparticipants, Cooperators, and Defectors might have netted nothing from joint enterprises in the long-run (Sigmund, 2010). The co-existence of negative emotions serves to instantiate the four strategies that yield cooperation via voluntary participation and altruistic punishment (Hauert et al., 2007; De Silva et al., 2010; Sigmund, 2010).

In the present study, interpretation of depression as an alternative strategy for nonparticipation was entitled the “nonparticipation hypothesis of depression”. The nonparticipation hypothesis appears feasible, both in the light of mathematical models for the emergence of cooperation and for the evolution, and in the light of supporting empirical observations. The hypothesis yields a functional explanation for the association between income inequality and depression. It also offers a potential mechanism for the observed social network-dynamics of emotion, and explains co-morbidity and proximal biological mechanisms for negative emotions. As predicted by the nonparticipation hypothesis, evidence about social dynamics superimposed with genetic variation does exist. It has been suggested, however, that most candidate-gene findings in psychiatry represent interplay of statistical errors and publication biases (Duncan and Keller, 2011). Although some of the above reviewed evidence may not hold due to such biases, the primary result derived herein concerns ‘gene’ mainly in the more historical meaning of the word; as a unit of inheritance

that may include arbitrary complex combinations of single-nucleotide polymorphisms (Plomin et al., 2009).

The precise evaluation of the nonparticipation hypothesis is necessarily a difficult task: it is difficult to draw clear borders between numerous joint enterprises that modern individuals are immersed in, or analyze their relative importance. It is also difficult to determine the amount and quality of alternative nonparticipation strategies. Yet, some dynamic model will be necessary for a satisfactory functional understanding of depression; the present suggestion is a simple one to start with as it incorporates only four-dimensional dynamics (four strategies). Higher-dimensional dynamical systems are increasingly difficult to analyze. It has been argued that “treating culture as a fixed variable seriously impedes our ability to understand and respond to disease states such as depression” (Kleinman, 2004). Herein, “treating culture as fixed variable” was avoided by examining individuals against a dynamic background of other individuals, although simplifications were done in order to track the ensuing complexity.

A number of research implications follow from adopting a combined evolutionary- and social-dynamics view point to depression. Regarding evolutionary dynamics it is difficult to separate adaptations that promote survival of a group from those that promote reproductive fitness of an individual. Herein, it was hypothesized that it is adaptive to retain in a population those traits that prevent adverse outcomes of imitation dynamics; namely, the evaporation of cooperation in important joint enterprises. Yet, evolutionary dynamics can also follow individual-based fitness. Careful analysis is needed to separate group-level effects from individual-level effects on one hand, and evolutionary dynamics from social-imitation dynamics on the other hand. Both evolutionary and imitation-based models tend to take similar mathematical forms (Nowak, 2006; Sigmund, 2010), but

obviously operate on different time-scales (multiple generations *versus* life-spans of individuals). This should justify the separation of time-scales for the model of this study.

The modeling results reviewed here pertained to a ‘well-mixed’ population where any individual has equal chances to meet any other individual. It has been shown, however, that network structures (distribution of social contacts between individuals) can play an important role in evolutionary and imitation-dynamical models (Nowak, 2006). Therefore, future studies evaluating the non-participation hypothesis may benefit from careful assessment of the structure of social network relevant to a given joint enterprise; typically, some network structures function as an amplifier of imitation/selection, whereas some network structures attenuate imitation/selection (Nowak, 2006). Regarding present results, however, it is reassuring that a recent empirical study concluded that “population structure has little relevance as a cooperation promoter or inhibitor among humans” (Grazia-Lázaro et al., 2012).

This study considered models where punishing of others is ‘costly’ for the individuals participating in a joint enterprise. Punishing need not be very costly always; for example, ostracism of defectors might be easy sometimes. Yet, people do engage in punishment practices that are costly, and such behaviors seem to be an important part of evolution. The partial nature of the driving social and evolutionary forces was explicitly modeled using the selection-strength parameters (s_i and s_g). Although non-costly punishment is an interesting research question in terms of modeling the evolution of cooperation (Hauert et al., 2007), it does not directly undermine any arguments presented here.

The nonparticipation hypothesis, in general, need not be an all-inclusive model of depression. Rather, it is intended to explain some functional aspects of depression etiology and evolutionary history. The hypothesis does not necessarily conflict with other theories of

depression. For example, psychic pain associated with low mood and sadness is sometimes seen to serve functions analogous to physical pain, but in response to social rather than physical adversity (Nesse, 1991; MacDonald and Leary, 2005; Hagen, 2011). Evolved abilities often provide more than one benefit for their carriers; for an extreme example, consider ability to grasp objects. Psychic pain hypothesis of depression does not appear exhaustive, however, as it faces difficulties in explaining suicidality and severe depression (Hagen, 2003, 2011; Friedman, 2012). The psychic pain of the depressed may also be interpreted within the nonparticipation hypothesis. Positive outcomes of various real-life joint enterprises frequently induce strong physiological responses and strong memories, which are not easily overridden by conscious reasoning; in order to have an effect on behavior, negative outcomes need comparable biological signals. Hence, psychic pain and pleasure may supplement or substitute the driving role of outcomes ('money') in some biologically relevant joint enterprises. Although 'money' is an intuitive conceptualization for a pay-off from joint enterprise, the models for evolutionary dynamics operate on deviations from average pay-off instead of any fixed units (Nowak, 2006; Sigmund, 2010). Humiliation due to failed defection or cooperation and unpleasantness/exhaustiveness of angry punishing may also serve to instantiate a biological 'pay-off structure'. As a learning signal, pain may motivate search for more profitable enterprises (starting from abstinence from current ones) and help to avoid unprofitable enterprises.

Depressive episodes are frequently associated with stressful life events (Kendler et al., 1998, 1999; Kendler and Gardner, 2010). The causal effect of stressful life events dependent on a person's own behavior, which are more strongly associated with depressive episodes than most other types of stressful life events, has been found modest, however (Kendler and Gardner, 2010). Furthermore, the increase in depressive-episode risk due to temporally well-isolated stressful life events tends to be short-lived, generally disappearing in

less than three months (Kendler et al., 1998; Surtees and Wainwright, 1999). Instead, the same study found that the stressful life events “associated with prolonged duration of risk—as exemplified by serious financial, legal, work, and marital problems—were usually more ‘difficulty-like’ in that they may have more ‘fuzzy’ onsets and are likely to persist some time” (Kendler et al., 1998). Such “difficulty-like” events can easily be subjectively interpreted as unfair social situations, and presence of defection, which is a causal antecedent of depression according to the present nonparticipation hypothesis.

A temporally isolated, impulse-like, stressful life event can plausibly function both as an impetus to re-evaluate one's behavioral strategy and as negative outcome of a joint enterprise. In addition to life outcomes, negative expectations have been linked with onset, maintenance, and biology of depressive episodes (Beck, 1967; Disner et al., 2011), and low moods appear socially contagious (Rosenquist et al., 2010). Elucidating precise relationships among the impulse-like stressful life events, expectations of adversity, and Public Good Games remains a future challenge. For example, one might attempt driving the evolution of strategies in a population directly through joint-enterprise outcomes (interpreted as ‘events’) as opposed to expected outcomes of strategies within the social reference group (‘imitation’ of high achievers), and thereby attempt to theoretically contrast actual instantaneous adversities *versus* realistic expectations of adversity.

Clinical implication of the nonparticipation hypothesis is that social dynamics and social fairness deserve more attention as potential causal antecedents of depression. In addition, the hypothesis implies that purely chemical extinguishing of all depressive symptoms from a population (Kleinman, 2012b) may be a formidable task due to suggested evolutionary underpinnings of the condition; if achieved, it might bring about some undesired cooperation-evaporating side-effects for the wider society. For more than twenty years, some

researchers have wondered “whether widespread use of antidepressants might be ... tampering with the mechanisms that regulate human social hierarchies” (Nesse, 1991).

In summary, the present modeling study, together with the reviewed previous research reports, suggest that depression offers a nonparticipation strategy to joint enterprises that are otherwise difficult to avoid, and that past evolutionary selection among groups of individuals, bands, or tribes, has favored this behavioral mode. This “nonparticipation hypothesis” serves to explain several seemingly unrelated empirical observations, such as association between income inequality and depression across populations, co-morbidity and proximal biological mechanisms for negative emotions, mechanisms for the social network-dynamics of emotion, and existence of a gene that interacts with environment in order to produce depression vulnerability in some people, thereby ensuring the potential for a nonparticipation response to defection in joint enterprises.

Acknowledgements

This study was financially supported by the Alfred Kordelin Foundation and the Academy of Finland (from the grant #258711); the sponsors had no role in manuscript preparation or decision to submit.

References

- Ahern, J., Galea, S., 2006. Social context and depression after a disaster: The role of income inequality. *J. Epidemiol. Community Health* 60, 766-770.
- Andrews, P.W., Thomson Jr, J.A., Amstadter, A., Neale, M.C. 2012. Primum non nocera: an evolutionary analysis whether antidepressants do more harm than good. *Front. Psychol.* 3, 117. doi: 10.3389/fpsyq.2012.00117
- Andrews, P.W., Thomson Jr, J.A., 2009. The bright side of being blue: Depression as an adaptation for analyzing complex problems. *Psychol. Rev.* 116, 620-654.
- Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. *Science* 211, 1390-1396.
- Beck, A.T., 1967. *Depression: Clinical, Experimental, and Theoretical Aspects*. University of Pennsylvania Press, Pennsylvania, Philadelphia.
- Benden, A., Farvolden, P., 2010. Depression and the workplace: A progress report. *Cur. Psychiat. Reports* 101, 73-79.
- Bernstein, J., Mishel, L., 1997. Has wage inequality stopped growing? *Mon. Labor Rev.* 3-16.
- Bienvenu, O.J., Davydow, D.S., Kendler, K.S., 2011. Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence. *Psychol. Med.* 41, 33-40.

- Boswell, P.C., Murray, E.J., 1981. Depression, schizophrenia, and social attraction. *J. Consult. Clin. Psychol.* 49, 641-647.
- Brosnan, S.F., De Waal, F.B.M., 2003. Monkeys reject unequal pay. *Nature* 425, 297-299.
- Brown, G.W., Harris, T.O., 2008. Depression and the serotonin transporter 5-HTTLPR polymorphism: A review and a hypothesis concerning gene–environment interaction. *J. Affect. Disord.* 111, 1-12.
- Buss, D.M., 2008. Human nature and individual differences: Evolution of human personality. In John, O.P., Robins, R.W., Pervin, L.A. (Eds.), *Handbook of personality: Theory and research* (3rd ed.). The Guildford Press, New York.
- Buss, D.M., Haselton, M.G., Shackelford, T.K., Bleske, A.L., Wakefield, J.C., 1998. Adaptations, exaptations, and spandrels. *Am. Psychol.* 53, 533-548.
- Carver, C.S., Miller, C.J., 2006. Relations of serotonin function to personality: Current views and a key methodological issue. *Psychiatry Res.* 144, 1-15.
- Caspi, A., Hariri, A.R., Holmes, A., Uher, R., Moffitt, T.E., 2010. Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *Am. J. Psychiatry* 167, 509-527.
- Clark, R., 2011. World income inequality in the global era: New estimates, 1990-2008. *Soc. Problems* 58, 565-592.
- Coleman, A.M., 2006. The puzzle of cooperation. *Nature* 440, 744-745.
- Crockett, M.J., 2009. The neurochemistry of fairness. *Ann. NY Acad. Sci.* 1167, 76-86.

- Crockett, M.J., Clark, L., Hauser, M.D., Robbins, T.W., 2010. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc. Natl. Acad. Sci. USA* 107, 17433-17438.
- Crockett, M.J., Clark, L., Tabibnia, G., Lieberman, M.D., Robbins, T.W., 2008. Serotonin modulates behavioral reactions to unfairness. *Science* 320, 1739.
- Coupland, C., Dhiman, P., Morriss, R., Arthur, A., Barton, G., Hippisley-Cox, J., 2011. Antidepressant use and risk of adverse outcomes in older people: population based cohort study. *Br. Med. J.* 343, d4551. doi: 10.1136/bmj.d4551
- Coyne, J.C., 1976. Depression and the response of others. *J. Abnorm. Psychol.* 85, 186-193.
- Darwin, C., 1871. *The Descent of Man, and Selection in Relation to Sex*. John Murray, London.
- De Silva, H., Hauert, C., Traulsen, A., & Sigmund, K., 2010. Freedom, enforcement, and the social dilemma of strong altruism. *J. Evolutionary Econ.* 20, 203-217.
- Disner, S.G., Beevers, C.G., Haigh, E.A.P., Beck, A.T., 2011. Neural mechanisms of the cognitive model of depression. *Nat. Rev. Neurosci.* 12, 467-477.
- Duncan, L.E., Keller, M.C., 2011. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* 168, 1041-1049.
- Farmer, A., Mahmood, A., Redman, K., Harris, T., Sadler, S., McGuffin, P., 2003. A sib-pair study of the Temperament and Character Inventory scales in major depression. *Arch. Gen. Psychiatry* 60, 490-496.

- Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 90, 980-994.
- Fiscella, K., Franks, P., 2000. Individual income, income inequality, health, and mortality: What are the relationships? *Health Serv. Res.* 35, 307-318.
- Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R.C., Fawcett, J., 2010. Antidepressant drug effects and depression severity. *JAMA* 303, 47-53.
- Fowler, J.H., 2005. Altruistic punishment and the origin of cooperation. *Proc. Natl. Acad. Sci. USA* 102, 7047-7049.
- Fowler, J.H., Christakis, N.A., 2008. The dynamic spread of happiness in a large social network. *BMJ* 337, a2338.
- Fowler, J.H., Christakis, N.A., 2010. Cooperative behavior cascades in human social networks. *Proc. Natl. Acad. Sci. USA* 107, 5334-5338.
- Friedman, R.A., 2012, January 17. Essay: Depression defies the rush to find an evolutionary upside. *The New York Times*, pp. D1.
- Gilbert, P., 2006. Evolution and depression: Issues and implications. *Psychol. Med.* 36, 287-297.
- Godin, I., Kittel, F., Coppieters, Y., Siegrist, J., 2005. A prospective study of cumulative job stress in relation to mental health. *BMC Public Health* 5, 67.
- Godoy, R.A., Reyes-García, V., McDade, T., Huanca, T., Leonard, W.R., Tanner, S., Vadez, V., 2006. Does village inequality in modern income harm the psyche? Anger, fear,

sadness, and alcohol consumption in a pre-industrial society. *Soc. Sci. Med.* 63, 359-372.

Gracia-Lazáro, C., Ferrer, A., Ruiz, G., Tarancón, A., Cuesta, J.A., Sánchez, A., & Moreno, Y., 2012. Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma. *Proc. Natl. Acad. Sci. USA* 109, 12922-12926.

Gächter, S., Renner, E., & Sefton, M., 2008. The long-run benefits of punishment. *Science* 322, 1510.

Hagen, E.H., 1999. The functions of postpartum depression. *Evol. Hum. Behav.* 20, 325-359.

Hagen, E.H., 2003. The Bargaining Model of depression, in: Hammerstein, P. (Ed.) *Genetic and Cultural Evolution of Cooperation*. The MIT Press, Cambridge, pp. 97-123.

Hagen, E.H., 2011. Evolutionary theories of depression: a critical review. *Can. J. Psychiatry*, 56, 716-726.

Hankin, B.L., Fraley, R.C., Lahey, B.B., Waldman, I.D., 2005. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. *J. Abnorm. Psychol.* 114, 96-108.

Hansenne, M., Reggers, J., Pinto, E., Kjiri, K., Ajamier, A., Ansseau, M., 1999. Temperament and Character Inventory (TCI) and depression. *J. Psychiatric Res.* 33, 31-36.

Harlé, K.M., Allen, J.J.B., Sanfey, A.G., 2010. The impact of depression on social economic decision making. *J. Abnorm. Psychol.* 119, 440-446.

- Harlé, K.M., Sanfey, A.G., 2007. Incidental sadness biases social economic decisions in the ultimatum game. *Emotion* 7, 876-881.
- Hauert, C., Traulsen, A., Brandt, H., Nowak, M.A., Sigmund, K., 2007. Via freedom to coercion: The emergence of costly punishment. *Science* 316, 1905-1907.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., Ziker, J., 2006. Costly punishment across human societies. *Science* 312, 1767-1771.
- Hill, A.L., Rand, D.G., Nowak, M.A., Christakis, N.A., 2010. Emotions as infectious diseases in a large social network: The SISa model. *Proc. Royal Society B: Biological Sci.* 277, 3827-3835.
- Hyde, J.S., Mezulis, A.H., Abramson, L.Y., 2008. The ABCs of depression : integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychol. Rev.* 115, 291–313.
- Jokela, M., Singh-Manoux, A., Shipley, M.J., Ferrie, J.E., Gimeno, D., Akbaraly, T. N., Head, J., Elovainio, M., Marmot, M.G., Kivimäki, M., 2011. Natural course of recurrent psychological distress in adulthood. *J. Affect. Disord.* 130, 454-461.
- Jylhä, P., Isometsä, E., 2006. Temperament, character and symptoms of anxiety and depression in the general population. *Eur. Psychiatry* 21, 389-395.
- Karg, K., Burmeister, M., Shedden, K., Sen, S., 2011. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited. *Arch. Gen. Psychiatry* 68, 444-454.

- Karpansalo, M., Kauhanen, J., Lakka, T.A., Manninen, P., Kaplan, G.A., Salonen, J.T., 2005. Depression and early retirement: Prospective population based study in middle aged men. *J. Epidemiol. Community Health* 59, 70-74.
- Keltikangas-Järvinen, L., Jokela, M., 2010. Nature and nurture in personality. *Focus* 8, 180-186.
- Kendler, K.S., Karkowski, L.M., Prescott, C.A., 1998. Stressful life events and major depression: Risk period, long-term contextual threat, and diagnostic specificity. *J. Nerv. Ment. Dis.* 186, 661-669.
- Kendler, K.S., Karkowski, L.M., Prescott, C.A., 1999. Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry*, 156, 837-841.
- Kendler, K.S., Gardner, C.O., 2010. Dependent stressful life events and prior depressive episodes in the prediction of major depression: the problem of causal inference in psychiatric epidemiology. *Arch. Gen. Psychiatry*, 67, 1120-1127.
- Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P. S., 2003. The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (NCS-R). *JAMA* 289, 3095-3105.
- Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T., 2008. Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 5, e45.
- Kleinman, A., 2004. Culture and depression. *N. Engl. J. Med.* 351, 951-953.

- Kleinman, A., 2012a. Rebalancing academic psychiatry: why it needs to happen—and soon. *Br. J. Psychiatry* 201, 421-422.
- Kleinman, A., 2012b. Culture, bereavement, and psychiatry. *Lancet* 379, 608-609.
- Lester, D., Gatto, J.L., 1990. Interpersonal trust, depression, and suicidal ideation in teenagers. *Psychol. Reports* 67, 786-786.
- Lee, D.T.S., Kleinman, J., Kleinman, A., 2007. Rethinking Depression: an ethnographic study of the experiences of depression among chinese. *Harv. Rev. Psychiatry*, 15, 1-8.
- MacDonald, G., Leary, M.R., 2005. Why does social exclusion hurt? The relationship between social and physical pain. *Psychol. Bull.* 131, 202-223.
- Marijnissen, G., Tuinier, S., Sijben, A.E.S., Verhoeven, W.M.A., 2002. The Temperament and Character Inventory in major depression. *J. Affect. Disord.* 70, 219-223.
- Mathers, C.D., Loncar, D., 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 3, e442.
- Merjonen, P., Keltikangas-Järvinen, L., Jokela, M., Seppälä, I., Lyytikäinen, L., Pulkki-Råback, L., Kivimäki, M., Elovainio, M., Kettunen, J., Ripatti, S., 2011. Hostility in adolescents and adults: A genome-wide association study of the young Finns. *Transl. Psychiatry* 1, e11.
- Muramatsu, N., 2003. County-level income inequality and depression among older americans. *Health Serv. Res.* 38, 1863-1884.
- Murray, C.J.L., Lopez, A.D., 1997. Alternative projections of mortality and disability by cause 1990-2020: Global burden of disease study. *The Lancet* 349, 1498-1504.

- Nesse, R.M., 1991. What good is feeling bad? The evolutionary benefits of psychic pain. *Sciences*, 31, 30-37.
- Nesse, R.M., 2000. Is depression an adaptation? *Arch. Gen. Psychiatry* 57, 14-20. Netterstrøm, B., Conrad, N., Bech, P., Fink, P., Olsen, O., Rugulies, R., Stansfeld, S., 2008. The relation between work-related psychosocial factors and the development of depression. *Epidemiol. Rev.* 30, 118-132.
- Nettle, D., 2004. Evolutionary origins of depression: A review and reformulation. *J. Affect. Disord.* 81, 91-102.
- Nowak, M.A., 2006. *Evolutionary dynamics: Exploring the equations of life*. The Belknap Press of Harvard University Press, Massachusetts, Cambridge.
- Nowak, M.A., 2012. Evolving cooperation. *J. Theor. Biol.* 299, 1-8.
- Olfson, M., Marcus, S.C., Druss, B., Elinson, L., Tanielian, T., Pincus, H.A., 2002. National trends in the outpatient treatment of depression. *JAMA* 287, 203-209.
- Pickett, K.E., James, O.W., Wilkinson, R.G., 2006. Income inequality and the prevalence of mental illness: A preliminary international analysis. *J. Epidemiol. Community Health*, 60, 646-647.
- Plomin, R., Haworth, C.M.A., Davis, O.S.P., 2009. Common disorders are quantitative traits. *Nat. Rev. Genet.* 10, 872-878.
- Price, J.S., Gardner, R., & Erickson, M., 2004. Can depression, anxiety, and somatization be understood as appeasement displays? *J. Affect. Disord.* 79, 1-11.

- Rosenquist, J.N., Fowler, J.H., Christakis, N.A., 2010. Social network determinants of depression. *Mol. Psychiatry* 16, 273-281.
- Rosenström, T., Jokela, M., Hintsanen, M., Pulkki-Råback, L., Hutri-Kähönen, N., Keltikangas-Järvinen, L., 2013. Longitudinal course of depressive symptoms in adulthood: linear stochastic differential equation modeling. *Psychol. Med.* 43, 933-944.
- Sigmund, K., 2010. *The calculus of selfishness*. Princeton University Press, New Jersey.
- Sigmund, K., De Silva, H., Traulsen, A., Hauert, C., 2010. Social learning promotes institutions for governing the commons. *Nature* 466, 861-863.
- Sloman, L., 2008. A new comprehensive evolutionary model of depression and anxiety. *J. Affect. Disord.* 106, 219-228.
- Solomon, A., Haaga, D.A.F., Arnow, B.A., 2001. Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research. *J. Nerv. Ment. Dis.* 189, 498-506.
- Sturm, R., Gresenz, C.R., 2002. Relations of income inequality and family income to chronic medical conditions and mental health disorders: National survey. *BMJ*, 324, 1-5.
- Sullivan, P.F., Neale, M.C., Kendler, K.S., 2000. Genetic epidemiology of major depression: Review and meta-analysis. *Am. J. Psychiatry* 157, 1552-1562.
- Surbey, M.K., 2011. Adaptive significance of low levels of self-deception and cooperation in depression. *Evol. Hum. Behav.* 32, 29-40.

- Surbey, M.K., McNally, J.J., 1997. Self-deception as a mediator of cooperation and defection in varying social contexts described in the iterated prisoner's dilemma. *Evol. Hum. Behav.* 18, 417-435.
- Surbey, M.K., Simpson, J., 2010. Cheerful cheaters beware! Presentation at The Human Behavior and Evolution Society Annual Meeting. Abstract:
http://www.hbes.com/conference/pdf/conference_22.pdf, Last accessed: 11.1.2013.
- Surtees, P.G., Wainwright, N.W.J., 1999. Surviving adversity: Event decay, vulnerability and the onset of anxiety and depressive disorder. *Eur. Arch. Psychiatry Clin. Neurosci.* 249, 86-95.
- Vöhringer, P.A., Ghaemi, S.N., 2011. Solving the antidepressant efficacy question: Effect sizes in major depressive disorder. *Clin. Ther.* B61.
- Watson, P.J., Andrews, P.W., 2002. Toward a revised evolutionary adaptationist analysis of depression: The social navigation hypothesis. *J. Affect. Disord.* 72, 1-14.
- Weich, S., Lewis, G., Jenkins, S.P., 2001. Income inequality and the prevalence of common mental disorders in Britain. *BMJ* 178, 222-227.
- Weissman, M.M., Bland, R.C., Canino, G.J., Faravelli, C., Greenwald, S., Hwu, H.G., Joyce, P.R., Karam, E.G., Lee, C.K., Lellouch, J., Lepine, J., Newman, S.C., Rubio-Stibec, M., Wells, E., Wickramaratne, P.J., Wittchen, H., Yeh, E., 1996. Cross-national epidemiology of major depression and bipolar disorder. *JAMA* 276, 293-299.
- Ylipaavalniemi, J., Kivimäki, M., Elovainio, M., Virtanen, M., Keltikangas-Järvinen, L., Vahtera, J., 2005. Psychosocial work characteristics and incidence of newly diagnosed

depression: A prospective cohort study of three different models. Soc. Sci. Med. 61, 111-122.

Accepted manuscript

Figure captions

Figure 1. Average system states (stationary distribution, or relative time in homogenous states of one strategy) as a function of selection strength (s_i) and available strategies (panel legends). The lower-right panel shows the average income of the modeled joint enterprise, or Public Good Game, for all three settings of available strategies (see other panels). Model parameters were set to equal those in the previous study by Hauert and others (2007); that is, $M = 100$, $r = 3$, $c = 1$, $\sigma = 1$, $\gamma = 0.3$, $\beta = 1$, and $N = 5$.

Figure 2. Time-dynamics of a joint enterprise in an infinite population with only three strategies available: defecting ('y-axis' in right-handed coordinates), cooperation (z-axis), and not participating (x-axis). All possible configurations for the proportions of strategies applied in the population are contained in a simplex (front triangle of the geometric illustration); corners describe a population with only one strategy, whereas the middle-section corresponds to the population where all the strategies are present in equal amounts. The arrows in the figure show the direction of more lucrative strategies for an individual situated in a population with the amounts of strategies indicated by the base of an arrow. The exact time-dynamics of the joint enterprise depend on the pay-off structure of the enterprise, but all enterprises that are (potentially) more beneficial than Nonparticipation show similar oscillating 'Rock-Paper-Scissors' dynamics: depending on the initial condition, state of the population follows some closed orbit (thick lines provide examples) on the direction determined by the gradient field (arrows). The time-average for pay-off of the joint enterprise is always equal to that gained without participating; due to incentive to defect, the joint enterprise does not benefit the population in the long-term unless additional strategies are

available. If the gain of the enterprise is small (the parameter r satisfies $1 < r \leq 2$), instead of oscillations, all possible trajectories eventually lead to Nonparticipants taking over the population. The figure was produced using equations by Sigmund (2010), with parameter-values $r=3$, $c=1$, $\sigma=1$, and $N=5$; as in the finite population model, and in other figures.

Figure 3. Probability of the system state with a depression gene, or a nonparticipation strategy (stationary distribution of the state). Probability of depression gene is shown as a function of imitation-selection strength s_i , genetic-selection strength s_g , number of groups (“#Groups”; columns of panels), and individuals per group (“Group size”; rows of panels). Other model parameters were set to equal those in Figures 1 and 2, and in the previous study by Hauert and others (2007); that is, $r = 3$, $c = 1$, $\sigma = 1$, $\gamma = 0.3$, $\beta = 1$, and $N = 5$.

Highlight bullets for the manuscript “Bargaining Models of Depression and Evolution of Cooperation”

A new model for evolutionary origins of depression

Depressed withhold benefits from others by not participating to joint enterprises

Multi-level selection among groups and social-strategies favor depression gene(s)

The model explains many seemingly unrelated empirical observations about depression

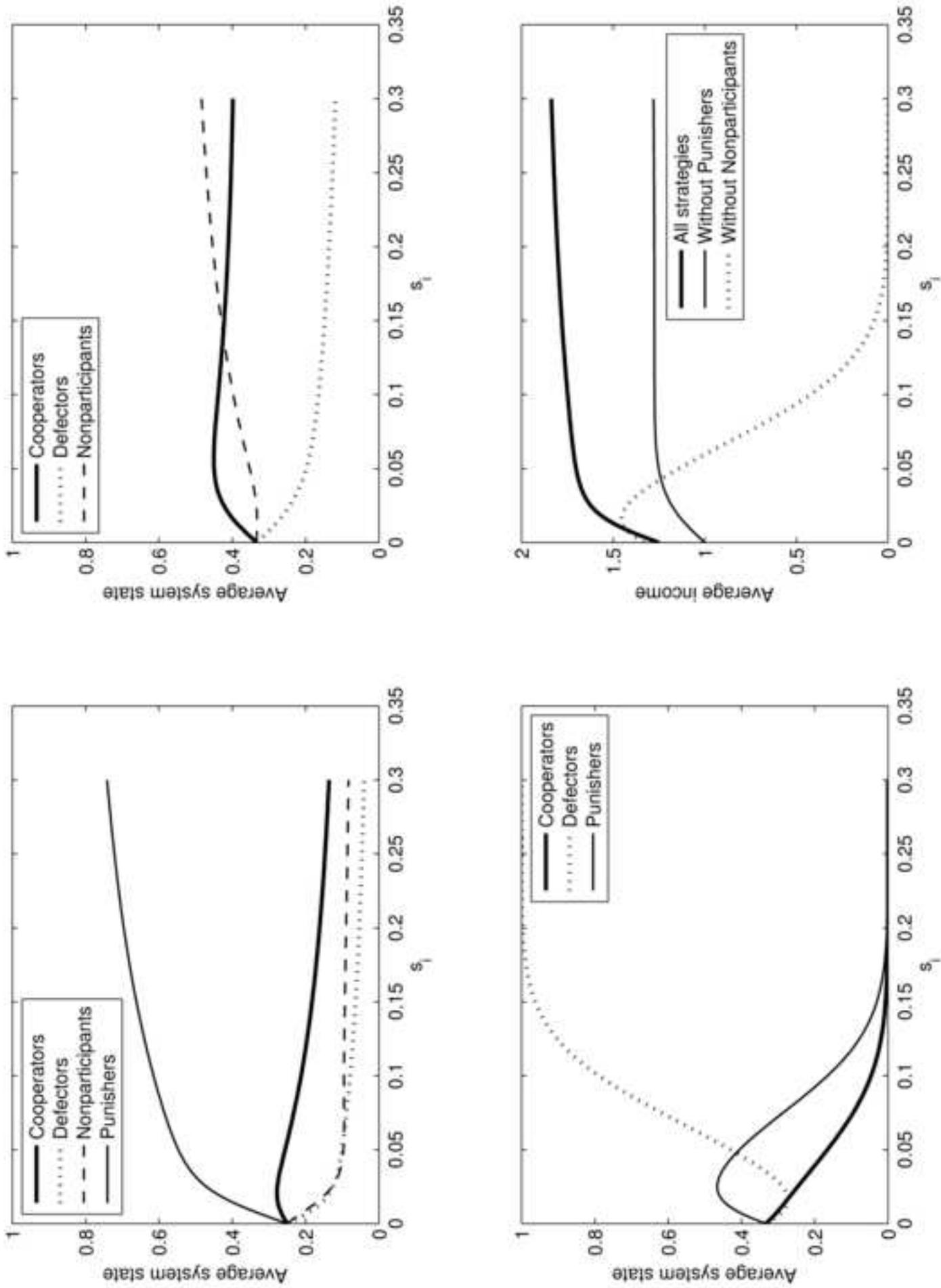
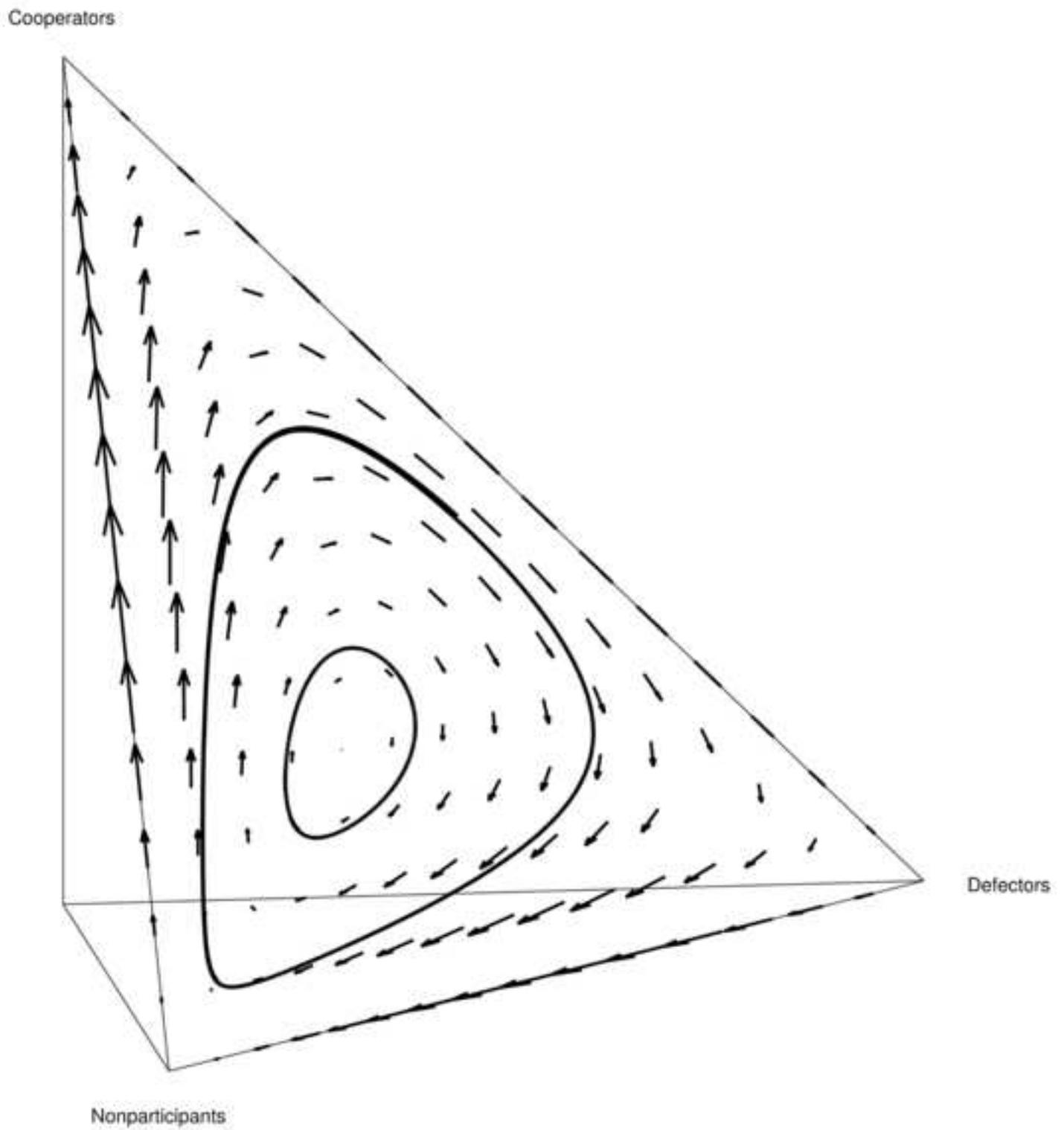


Figure 1



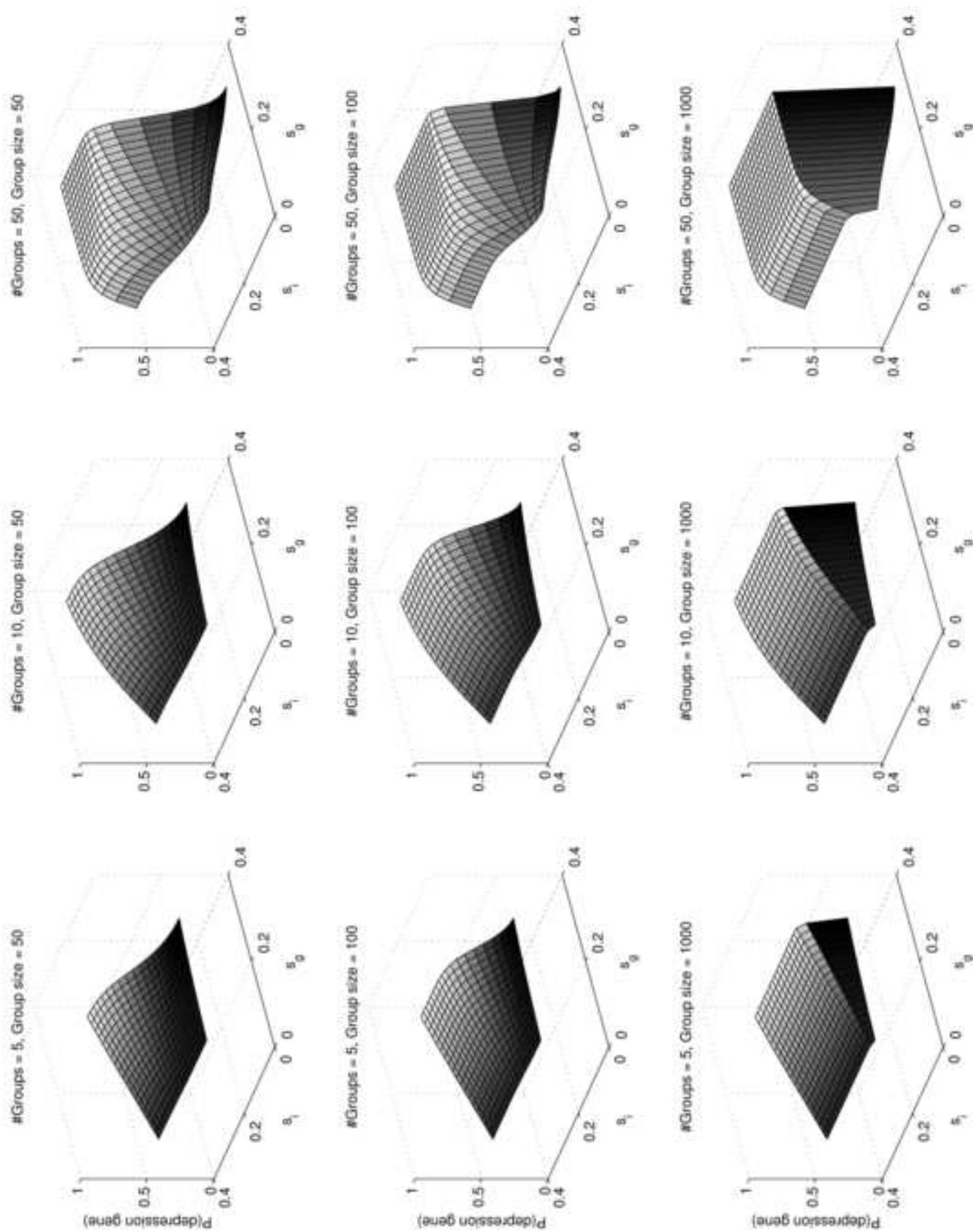


Figure 3