Title: A new agent-based model offers insight into population-wide adoption of prosocial common-pool behavior

Authors: Garry Sotnik, Thaddeus T. Shannon, & Wayne W. Wakeland

Abstract: Prosocial common-pool behavior is a subtractable and non-excludable resource that benefits those exposed. While such behavior is ubiquitous, individuals who adopt it are usually at a disadvantage to freeriders - those who benefit from it without adopting. How populations maintain or increase adoption of such behavior is unclear and remains an active research question. Our objective was to explore the influence of important factors on population-wide adoption of prosocial common-pool behavior. We developed a new theoretical model, called Multilevel Group Selection I, that establishes previously unexplored relationships among group composition and size, movement between groups, multilevel group selection, population density, pressure to change behavior, and synergy from prosocial common-pool behavior. We first conducted analysis of agent behavior and model dynamics under alternative parameter combinations. We then developed a variant of the model that sweeps through a preset range of parameter combinations and used it to explore simulation outcomes. Groups with larger numbers of prosocial agents are better able to withstand pressure than groups with lower numbers. High population density prevents population-wide adoption. Movement between groups assists but is not required for population-wide adoption. Different pressure-synergy combinations can produce substantially different adoption outcomes (extinction, persistence, and full adoption). Multilevel group selection is necessary for population-wide adoption. Population-wide adoption of prosocial common-pool behavior only occurs under a limited set of parameter combinations. The model offers a surprisingly large degree of behavioral richness, given its simple structure and rules, and serves as a fertile base for a fruitful and socially-relevant research program.

Keywords: agent-based model, common-pool resource, freeriding, multilevel group selection, prosocial common-pool behavior, social change.

1 Introduction

We live in a rapidly changing world and adopting new behavior is vital for both individual and collective benefit. Evidence shows, however, that many individuals at least initially do not adopt new behavior (Juhola et al., 2016). There are a number of potential explanations for why. One is that they may have a general skepticism or aversion to change (Rogers, 1983). Another is that they may perceive the new behavior as a fad (Abrahamson, 1991) or a maladaptation (Barnett & O'Neill, 2010; Magnan et al., 2016). Yet another explanation, which is the focus of this paper, is that there is a cost to adopting the new behavior that they may want to avoid or at least delay incurring. The size of this cost associated with adopting the new behavior is often influenced by an individual's (intellectual, financial, physical) ability to provide the behavior, their need for it, and their benefit from it.

The last explanation for why an individual might not adopt new behavior is especially probable when the behavior is prosocial (i.e., behavior like helping, sharing, or cooperating that benefits others [Twenge et al., 2007]). In such cases, adoption of the new behavior by others produces positive externalities (i.e., positive side effects that benefit those exposed [Cornes & Sandler, 1996]), which provide exposed non-adopters with an opportunity and incentive to freeride (i.e., benefit from the behavior without adopting it [Battaglini et al., 2012; Kim & Walker, 1984; Ozono et al., 2017]). Therefore, the ability to freeride likewise influences the associated cost and whether an individual adopts. When the behavior is also a common-pool resource, i.e., is both subtractable (the benefit from the behavior by one individual reduces possible benefit by another) and non-excludable (the benefit from the behavior is not easily limited to specific individuals) (Ostrom et al., 1994), freeriding by others increases the cost and reduces the benefit of the new behavior to existing and potential adopters, in turn, further discouraging its adoption and reducing its presence in a population.

To date, much of the theoretical progress in relation to prosocial common-pool behavior (e.g., cooperation [Axelrod, 1984]) has been made in relation to two individuals facing a situation with a cost-benefit structure that is commonly referred to as the Prisoner's Dilemma (von Neumann & Morgenstern, 1944). An example of this progress is the identification of the tit-for-tat strategy (Axelrod & Hamilton, 1981; Dal Bó & Fréchette, 2019) that helps those implementing prosocial common-pool behavior in such a situation avoid loss in interaction with those who are determined to exploit them by not reciprocating. However, many (if not most) of the situations that individuals face as they interact with others are not of the Prisoner's-Dilemma type (e.g., bargaining, coalition building [Maskin, 2016]). Additionally, many groups that play important roles in society are composed of more than two individuals, including (offline and online) communities and classrooms, friendships, households, and professional organizations. Therefore, developing theory of prosocial common-pool behavior for situations in groups of size greater than two is equally if not more important than developing it for situations in groups of size two. However, simply extending the Prisoner's Dilemma to groups of size greater than two by either coupling group members into two sub-groups (e.g., Camerer, 2003; von Neumann & Morgenstern, 1944) or engaging group members in pair-wise (as opposed to multi-wise) interaction (e.g., Helbing et al., 2011; Helbing & Yu, 2008; Nowak & May, 1992; Nowak & Sigmund, 1998; Tarnita et al., 2009) ignores the common-pool nature of such groups, i.e., it overlooks the effect group size and composition have on a situation's cost-benefit structure (more on this in Sections 3.1 and 4.1.1).

Furthermore, many groups overlap as individuals hold membership in more than one group. Each group provides its members with an opportunity to be a provider or a recipient of prosocial common-pool behavior, as well as an opportunity to freeride or reciprocate. Experience from interaction within any one of the groups also influences whether an individual repeats or adopts the same or other prosocial common-pool behavior in that and/or other groups. The relative proportion of members providing prosocial common-pool behavior in a group is also likely to make the cost-benefit structure in some groups more favorable than in others (Boyd, 2018; Goodnight & Stevens, 1997; Sober & Wilson, 1999; Waring et al., 2015; Wilson & Sober, 1994; Wilson & Wilson, 2007). Such between-group interaction (e.g., Bowles & Gintis, 2009; Boyd & Richerson, 1992; De Silva & Sigmund, 2009; Frank, 2009; Hauert et al., 2002, 2007; Wang et al., 2011; Ye et al., 2011) is as important to a theory of prosocial common-pool behavior as the aforementioned within-group interaction (exemplified by the Prisoner's Dilemma). Yet, current theory on cooperative behavior does not address it (Maskin, 2016).

Therefore, there is a need to develop theory for prosocial common-pool behavior in groups of size greater than two; under different levels of cost, benefit, and need; and taking into account freeriding and differences in group composition. Such theory has the potential to help with designing mechanisms aimed at preventing the exploitation of prosocial common-pool behavior and facilitating its adoption. It would also have the potential to provide insight into the relationship between prosocial common-pool behavior and tangible common-pool resources (more on this in Sections 2 and 6).

To this end, we developed a new theoretical model in NetLogo (version 6.2.0; Wilensky, 1999), called Multilevel Group Selection I (version 2.0; henceforth “MGS I” or “the base model”) to explore the influence of important factors on population-wide adoption of prosocial common-pool behavior. Our long-term aim in developing the model was to create a fertile base upon which more complex models - its variants - aimed at studying population-wide adoption of prosocial common-pool behavior can be built. We then developed a variant of the model that sweeps through a preset range of parameter combinations and used it to explore simulation outcomes.

In this paper, we first define prosocial common-pool behavior in Section 2. In Section 3, we introduce the new MGS I model and describe our process of verifying it. In Section 4, we describe our analysis of agent behavior and overall model dynamics. In Section 5, we present simulation results and use them to verify the expected model dynamics and to analyze the influence of population density, the initial percent of contributors, and movement. Finally, in Section 6, we draw some initial conclusions from the analysis, discuss their relevance for regulating exploitation of tangible common-pool resources, and share ideas for possible variants. The base model, its variant (MGS I Sweep), and the Python (version 3.6; van Rossum & Drake Jr., 2009) scripts used for analysis and visualization are available on GitHub: https://github.com/Multilevel-Group-Selection.

2 The concept of prosocial common-pool behavior

We define prosocial common-pool behavior as a subtractable resource that benefits those exposed, is provided by an individual through interaction with others, and the access to which is not easily limited (i.e., is non-excludable). Prosocial common-pool behavior is both similar and different from other types of common-pool resources, namely tangible (e.g., fisheries, forests, groundwater basins [Ostrom, 1999]) and intangible (e.g., information [Hess & Ostrom, 2003]). For example, both tangible common-pool resources and prosocial common-pool behavior can be depleted, i.e., they are both subtractable. However, in addition to being physically depletable (an individual engaging in prosocial common-pool behavior can become physically or mentally exhausted), a prosocial common-pool behavior can also be discouraged by social factors (such as who is the depletor), new information, or changes in other (internal or external) factors. In other words, unlike other types of common-pool resources, prosocial common-pool behavior has physical, psychological, and social dimensions, and, therefore, has both tangible and intangible characteristics. Prosocial common-pool behavior (subtractable and non-excludable) is also both similar and different from other types of behavior, namely club (non-subtractable and excludable), private (subtractable and excludable), and public (non-subtractable and non-excludable). For example, access to both public and common-pool behavior cannot be easily limited to specific individuals (i.e., they are both non-excludable). However, benefit from prosocial public behavior is not subtractable (e.g., introducing laws, planting trees), while benefit from prosocial common-pool behavior is (e.g., volunteering at a homeless shelter).

A clear example of prosocial common-pool behavior is collective voluntary labor, such as building and cleaning (e.g., Simon & Mobekk, 2019). A less obvious example is prosocial competition (e.g., Gilbert & Basran, 2019), which involves competing in a way that promotes a healthy environment for others and oneself. Then, of course, there is adhering to shared rules in consuming a tangible common-pool resource (e.g., Ostrom et al., 1994), which reduces the current depletion of a resource, thereby benefiting others and oneself by making more of the resource available for future consumption. Managing a tangible common-pool resource, therefore, requires, at least in part, managing the depletion of the corresponding prosocial common-pool behavior (more on this in Section 6.2).

Sources of motivation for prosocial common-pool behavior can vary substantially. Prosocial common-pool behavior might be motivated by altruism, which is driven by concern for others; egoism, which is driven by social status, reputation, or expectation for reciprocity; or by a desire to adhere to a personal ideology. As mentioned above, adoption of new prosocial common-pool behavior might be hindered by skepticism or aversion to change, perception of the new behavior as a fad or maladaptation, or a desire to avoid or delay an associated cost. A prosocial common-pool behavior can also provide immediate or future benefit to its provider, be either free to its provider or at a cost, it can be synergistic, and it can have a local or global, and a current or future, impact.

3 The Multilevel Group Selection I model

3.1 Description

Multilevel Group Selection I (version 2.0; MGS I) simulates a social space composed of agents that either contribute prosocial common-pool behavior when in a group or not (i.e., freeride), and that are equally exposed to some (physical, psychological, and/or social) pressure to change their behavior. Exposure to prosocial common-pool behavior can help agents withstand the pressure, especially when the behavior is synergistic. Agents move within their social space to avoid this pressure, as a result forming into groups (common pools), which is where prosocial common-pool behavior can occur. Agents unable to withstand the pressure automatically change from contributing prosocial common-pool behavior to not or vice versa. We leave the empowerment of agents with the ability to choose their behavior for future (more complex) variants of this base model.

We conceptualize the social space that the agents move within as the maximum social structure (network) of a population of agents. In our base model, this space takes the form of a square lattice, Z^2, with each square representing a social spot (node), $z_{ij} \in Z^2$, that can be empty or occupied by one agent. The lattice serves as a unique constraint on the structure of the population that for our purposes conveniently focuses the study of within- and between-group (cluster) interactions on groups ranging between 2 and 9 in size. We avoid boundary effects in the lattice by forming it into a torus. The social space's density, D, determines how many of the space's spots are occupied by agents during a simulation. Which spots are occupied is initially random and afterwards is determined by agent movement. Exploring alternative topologies is an opportunity for future variants.

There are N agents, labeled a_i, where i = 1, 2, … N and N is determined by the product of the social space's density and size, $D*Z^2$. Initially, agents are randomly distributed (using a uniform distribution) throughout the social space. Each agent's initial behavior is assigned randomly, using the initial percent of contributors, $I_{t=0}$. At the beginning of each iteration, each agent is also endowed with an equal level of effort, e_i, which in the base model we normalized across agents at 1. Agents need effort to provide prosocial common-pool behavior and withstand pressure. The effort sets a limit to how much prosocial common-pool behavior an agent can contribute in a group and there is a 1:1 relationship between the two. It takes all of an agent's effort to contribute prosocial common-pool behavior to a group. On the other hand, non-contributors retain all of their effort. We designed the base model and framed this study with humans in mind. However, we do not see a reason why the base model or a variant could not apply equally well in the context of other social life forms.

Groups represent all possible interactions among agents, which, in the real world, can vary substantially in purpose and many other characteristics (e.g., duration, norms). A group is defined from the perspective of its focal agent. A focal agent's group, $g_{i,t}$, at any point during a simulation is composed of its Moore neighborhood neighbors and, as agents move, can vary from 2 to 9 agents in size (including the focal agent) during a simulation. The dynamic number of groups at any point during a simulation, $g_{i,t} \in G_t$, equals the number of agents who have at least one neighbor, i.e., $G_t \leq N = D * Z^2$. Any agent that has a group is also in the groups of its neighbors (group members). In other words, the groups of neighboring agents overlap and, as a result, an agent can be in 1 to 8 other groups (in addition to its own), depending on whether it has 1 or 8 neighbors, respectively (Figure 1). An agent contributes (or not) an equal amount of effort in each one of its groups. An agent can also be outside of a group, i.e., alone, in a social spot with no other agents within its Moore neighborhood. An agent that is not in a group cannot contribute or benefit from prosocial common-pool behavior.

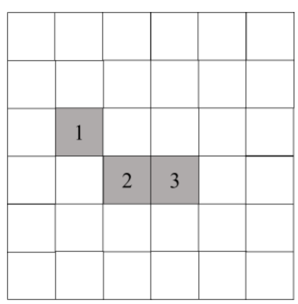

Figure 1: An example-based explanation of how groups are defined in MGS I. The 6x6 landscape includes three agents that are organized into three groups. Agent 1's group consists of agent 2 and itself. Agent 2's group consists of agent 1, 3, and itself. Agent 3's group consists of agent 2 and itself.

An agent is able to withstand the pressure to change its behavior when the sum of its retained effort and the benefit from its group is more than the pressure level. An agent can only derive benefit from its own group, i.e., the group within which it is the focal agent. There are two leading approaches to calculating the benefit from a group of a size greater than two. One is based on sequential pairwise interaction between an agent and its neighbors, from which the individual benefits to the agent are summed (e.g., Helbing et al., 2011; Helbing & Yu, 2008; Nowak & May, 1992; Nowak & Sigmund, 1998; Tarnita et al., 2009). This approach, however, does not account for the subtractable and non-excludable characteristics of an agent's prosocial common-pool behavior within a group. It does not account for the fact that the amount of effort available to an individual agent within a group decreases with every interaction with another agent within that group nor the fact that an agent's behavior within a group cannot be limited to one member at a time. Within the context of common-pool behavior, the approach inflates both how much benefit an agent can provide to and can receive from its group members, which biases agent behavior and potentially simulation results in favor of larger groups and prosocial agents.

The other leading approach to calculating benefit from groups of size greater than two is based on simultaneous multi-wise interaction among the agent and its neighbors, from which the individual benefit to the agent is averaged (e.g., Bowles & Gintis, 2009; Boyd & Richerson, 1992; Choi & Robertson, 2019; De Silva & Sigmund, 2009; Frank, 2009; Hauert et al., 2002, 2007; Wang et al., 2011; Ye et al., 2011). This approach, which is a key building block in our base model, reflects the subtractable and non-excludable characteristics of prosocial common-pool behavior within a group. The resulting ability of a contributor and non-contributor to withstand pressure to change its behavior are calculated as follows:

$$ability\ of\ contributor = s * \left(\sum_{j=1}^{N_i} e_j\right) / N_i \qquad \text{Equation 1}$$

$$ability\ of\ non-contributor = e_i + s * \left(\sum_{j=1}^{N_i} e_j\right) / N_i \qquad \text{Equation 2}$$

where e_i is the effort that each agent is endowed with at the beginning of each iteration and which, when in a group, a contributor contributes and a non-contributor keeps; s is the level of synergy from the prosocial common-pool behavior, and e_j is the effort contributed by a group member (including the focal agent, if it is a contributor). The ability of an agent that is not in a group to withstand the pressure to change its behavior is equal to the level of its unused effort, e_i, with which it was endowed at the beginning of the iteration and which it did not have an opportunity to contribute. As a result, the ability to withstand the pressure to change behavior is the same for all agents that are alone, but may differ for agents in groups, based on their type and group's composition. The amount of effort in possession resets for every agent each iteration, i.e., it does not accumulate.

An iteration during simulation of the base model consists of two main steps, potential movement and potential behavioral change. Agents can move within their social space, which represents any process of leaving and/or joining a group. It could represent physical movement from one geographic location to another or virtual movement from one online location to another. Agents are selected randomly one-by-one and automatically move to one of the closest empty spots if their ability to withstand pressure (the amount of effort in their possession) is below the pressure level. As is the case with changing behavior, we leave the empowerment of agents with the ability to choose where they move for future (more complex) variants of this base model. Having agents move to the closest empty spot is intended to reflect limitations of local knowledge and could be relaxed in future variants. After potentially moving, agents are selected randomly one-by-one and automatically change their behavior if their ability to withstand pressure (the amount of effort in their possession) is below the pressure level.

3.2 Verification

We took steps to verify that the base model correctly represented the conceptual model that we described in Section 3.1. The model verification involved the following three sets of activities: (a) code walkthroughs, and analyses of agent behavior under alternative (b) parameter settings and (c) module deactivations. The code walkthroughs involved discussions of every code line's fit with the corresponding components of and often led to simplifying changes to earlier versions of the conceptual model. The analyses of agent behavior were conducted with the model set in slow-mode and involved observing whether randomly selected agents behaved as was expected. The analysis under alternative parameter settings involved varying parameters over their full plausible ranges and observing results that were consistent with design objectives and analytically predicted end states. Module deactivations included suppressing movement and/or behavior change, and assuring the correctness of the results.

4 Theoretical analyses

We analyzed MGS I from a micro-level (agent behavior) and a macro-level (model dynamics) perspective. Despite the base model's relative simplicity (in comparison to the complexity of real-world social contexts and models that attempt to capture them), its outcomes vary substantially under alternative pressure-synergy combinations and, as we show herein, provide plausible explanations for real-world outcomes/interpretations.

4.1 Micro-level analysis of agent behavior

4.1.1 Link to game theory

Agents in MGS I can face situations that are commonly studied in game theory. As mentioned in the introduction, much of the theoretical progress in relation to prosocial common-pool behavior has been made in relation to two individuals facing a situation with a payoff structure that is commonly referred to as the Prisoner's Dilemma (Table 1).

Table 1: The payoff structure faced by individuals in a Prisoner's Dilemma, where T (temptation) > R (reward) > P (punishment) > S (sucker's).

	Cooperate	Defect
Cooperate	R, R	S, T
Defect	T, S	P, P

To demonstrate that agents in MGS I can face the Prisoner's Dilemma, we first extended the interpretation of the classic 2x2 payoff structure to groups of size greater than two. An agent's ability to withstand pressure to change its behavior (i.e., its payoff, in game theoretic terminology) equals: T (temptation) when it does not contribute prosocial common-pool behavior while at least one agent in its group does, R (reward) when it and all the other agents in its group contribute, P (punishment) when it and the other agents in its group do not contribute, and S (sucker's) when the agent contributes while at least one agent in its group does not. This extension is applicable to groups of all sizes.

We then identified the combinations of (group) size and synergy (from prosocial common-pool behavior) in which agents can face the Prisoner's Dilemma (Figure 2). We did the same for another (albeit less) popular situation in game theory, commonly referred to as the coordination game and defined by the following payoff structure: R > T and P > S (Figure 2). Interestingly, outside of the two extreme situations in which every group member either contributes or doesn't (and receives an R or P, respectively), how many of the group members contribute (the group's composition) does not affect whether they face a Prisoner's Dilemma (under the aforementioned size-synergy combinations). Contribution decisions do, however, affect whether or not agents face a coordination game situation, due to the resulting values of T and S.

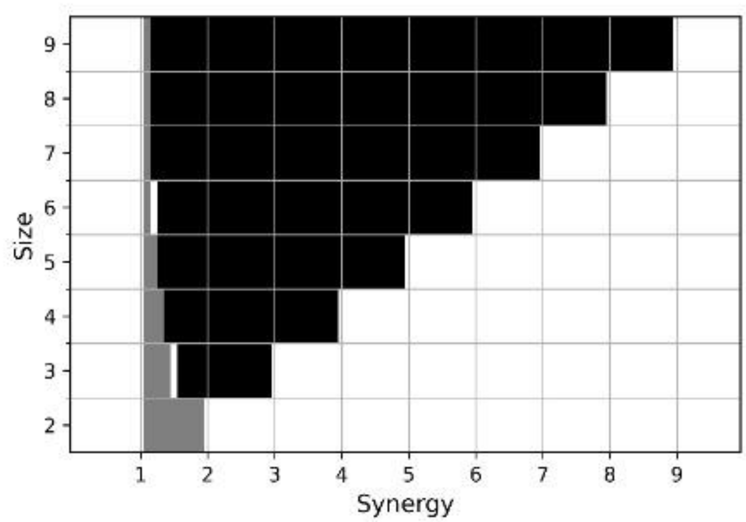

Figure 2: The (group) size-synergy (from prosocial common-pool behavior) combinations in which agents can face a Prisoner's-Dilemma-type (grey), a coordination-game-type (black), or some other type of a situation (white). The data underlying the figure was generated with Python's numpy package (version 1.20; Harris et al., 2020). The figure was generated with Python's matplotlib package (version 3.4.1; Hunter, 2007).

In addition to potentially facing the Prisoner's-Dilemma or the coordination-game-type of situations, MGS I agents can also find themselves in other types of situations. In fact, MGS I agents face neither of the two situations in more than half of the size-synergy combinations (Figure 2), suggesting a surprisingly large degree of behavioral richness in the model, despite its simple structure and rules.

4.1.2 The influence of group composition

Group composition has a substantial impact on an agent's ability to withstand the need to change its behavior and, in turn, on population-wide adoption of prosocial common-pool behavior. Table 2 illustrates this by listing the benefit to contributors and non-contributors from groups of different compositions. If an agent's benefit from a group is equal to or less than the pressure to change its behavior, the agent will need to change it.

Table 2: The benefit from groups with different combinations of agents from the perspective of (a) contributors and (b) non-contributors. The values in the tables are based on effort set to 1 and synergy from prosocial common-pool behavior set to 2.

(a)

# of contributors \ # of non-contributors	0	1	2	3	4	5	6	7	8
1	2.00	1.00	0.67	0.50	0.40	0.33	0.29	0.25	0.22
2	2.00	1.33	1.00	0.80	0.67	0.57	0.50	0.44	
3	2.00	1.50	1.20	1.00	0.86	0.75	0.67		
4	2.00	1.60	1.33	1.14	1.00	0.89			
5	2.00	1.67	1.43	1.25	1.11				
6	2.00	1.71	1.50	1.33					
7	2.00	1.75	1.56						
8	2.00	1.78							
9	2.00								

(b)

# of non-contributors \ # of contributors	0	1	2	3	4	5	6	7	8
1	1.00	2.00	2.33	2.50	2.60	2.67	2.71	2.75	2.78
2	1.00	1.67	2.00	2.20	2.33	2.43	2.50	2.56	
3	1.00	1.50	1.80	2.00	2.14	2.25	2.33		
4	1.00	1.40	1.67	1.86	2.00	2.11			
5	1.00	1.33	1.57	1.75	1.89				
6	1.00	1.29	1.50	1.67					
7	1.00	1.25	1.44						
8	1.00	1.22							
9	1.00								

Next, we used Equations 1 and 2 (which calculate an agent's ability to withstand pressure) to explore the relationship between group composition and pressure. We first defined a group's composition as the number of its contributing members divided by the total number of its members, $(\sum_{j=1}^{N_i} e_j) \,/\, N_i$. We then set the equations equal to pressure, since this is the point that determines if an agent needs to change its behavior. Finally, we expressed the equations from the perspective of group composition and through the pressure and synergy parameters:

$$Group\ composition\ of\ contributor = p/s \qquad \text{Equation 3}$$

$$Group\ composition\ of\ non-contributor = (p-e)/s \qquad \text{Equation 4}$$

Figure 3 illustrates the relationships expressed in Equations 3 and 4 at different levels of group composition and pressure to change behavior, and with the synergy from prosocial common-pool behavior set equal to 2. A contributor or non-contributor that faces a situation in which their composition-pressure combination is to the right of the orange or blue line, respectively, will need to change its behavior.

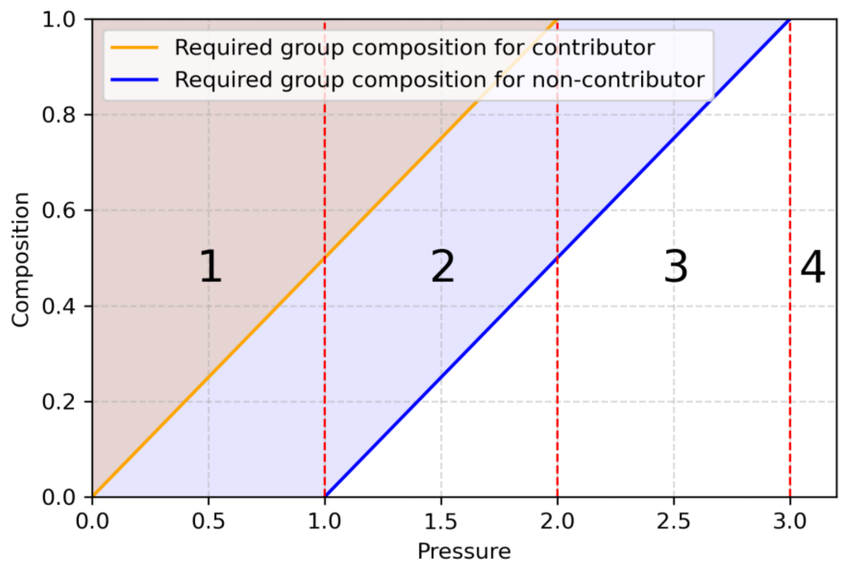

Figure 3: The relationship between group composition and pressure to change behavior, from the perspective of contributors and non-contributors. The data underlying the figure was generated with Python's numpy package (version 1.20; Harris et al., 2020). The figure was generated with Python's matplotlib package (version 3.4.1; Hunter, 2007).

As the broken red lines in Figure 3 indicate, composition-pressure combinations faced by agents can be partitioned into four segments, separated by three critical pressure points.

- The <u>1st segment</u> consists of composition-pressure combinations in which pressure is less than effort ($p < e = 1$). This is the satisfaction boundary for lone agents. A contributor might not be able to withstand pressure to change behavior if it faces a composition-pressure combination that falls to the right of the orange line. A non-contributor facing a combination from this segment does not need to change behavior.

- The 1st critical point is where pressure equals effort ($p = e = 1$). At this point and at higher pressure, non-contributors need prosocial group members to withstand pressure to change behavior.
- The 2nd segment consists of composition-pressure combinations in which pressure is equal to or greater than effort but less than synergy ($e \leq p < s$). Both a contributor and a non-contributor might not be able to withstand pressure to change behavior if it faces a composition-pressure combination that falls to the right of the orange or blue line, respectively.
- The 2nd critical point is where pressure equals synergy ($p = s = 2$). This is the satisfaction boundary for contributing agents in groups with 100% prosocial behavior. At this point and at higher pressure, contributors are not able to withstand pressure to change behavior.
- The 3rd segment consists of composition-pressure combinations in which pressure is equal to or greater than effort but less than the sum of effort and synergy from prosocial common-pool behavior ($e \leq p < e + s$). All contributors are not able to withstand pressure to change behavior, while a non-contributor is not able to if the combination it faces falls to the right of the blue line.
- The 3rd critical point is where pressure equals the sum of effort and synergy ($p = e + s$). At this point and at higher pressure, both contributors and non-contributors are not able to withstand pressure to change behavior.
- The 4th segment consists of composition-pressure combinations in which pressure is equal to or greater than the sum of effort and synergy ($e + s \leq p$). All contributors and non-contributors are not able to withstand pressure to change behavior.

Ceteris paribus, a change (increase/decrease) in the effort level shifts the blue line (right/left). It affects the non-contributors directly by increasing/decreasing their safe zone, and both contributors and non-contributors indirectly through the benefit they receive from their groups. An increase (decrease) in the value of prosocial common-pool behavior decreases (increases) the slope of the orange and blue line, thereby increasing the safe zones and the size of segments 2 and 3.

Similarly, Figure 4 illustrates the relationships expressed in Equations 3 and 4 with different group composition and synergy from prosocial common-pool behavior combinations, and this time with pressure (instead of synergy) set equal to 1.5. As in Figure 3, a contributor or non-contributor that faces a situation in which their composition-synergy combination is to the left of the orange or blue line, respectively, will need to change its behavior.

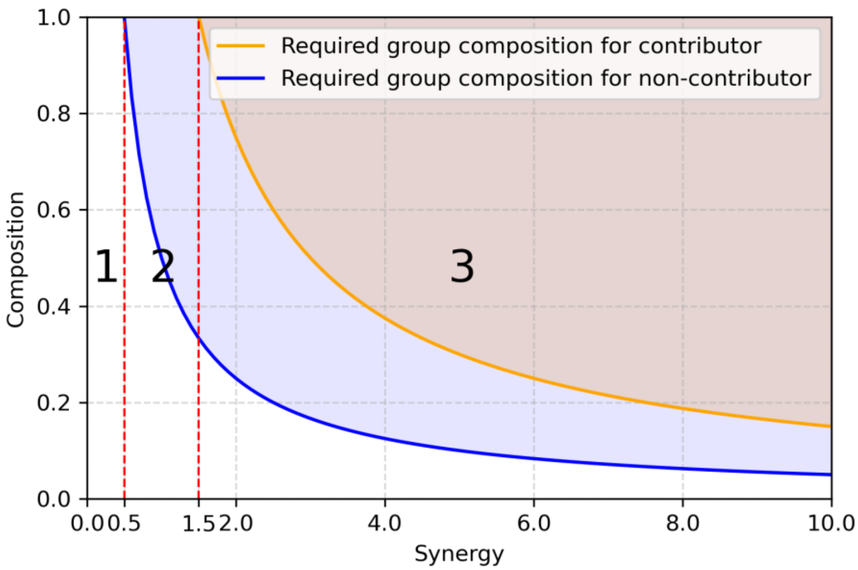

Figure 4: The relationship between group composition and synergy from prosocial common-pool behavior, from the perspective of contributors and non-contributors. The data underlying the figure was generated with Python's numpy package (version 1.20; Harris et al., 2020). The figure was generated with Python's matplotlib package (version 3.4.1; Hunter, 2007).

As the broken red lines in Figure 4 indicate, composition-synergy combinations faced by agents can be partitioned into three segments, separated by two critical value points.

- The 1st segment consists of composition-synergy combinations in which synergy is lower than the difference between pressure and effort, $s < p - e$. In such cases, both contributors and non-contributors are better off avoiding groups.
- The 1st critical point is equal to the difference between pressure and effort, $p - e$. At this point and at higher synergy, it becomes beneficial to not-contribute.
- The 2nd segment consists of composition-synergy combinations in which synergy is equal to or greater than the difference between pressure and effort, but is less than pressure, $p - e \leq s < p$. In such cases, non-contributors begin to benefit from groups with large numbers of contributors. On the other hand, contributors remain better off avoiding groups.

- The <u>2nd critical point</u> is where synergy equals pressure, $s = p$. At higher synergy levels, it becomes beneficial for contributors to join groups.
- The <u>3rd segment</u> consists of composition-synergy combinations in which synergy is greater than pressure, $p < s$. In such cases, both contributors and non-contributors benefit from groups with large numbers of contributors.

Ceteris paribus, a change (increase/decrease) in effort shifts (rightward/leftward) the first critical point. It affects the non-contributors directly by expanding/reducing their safe zone, and both contributors and non-contributors indirectly through the benefit they receive from their groups. An increase (decrease) in pressure reduces (increases) the concavity of the orange and blue lines and shifts them leftward, thereby reducing the safe zones.

Finally, in addition to influencing the cost-benefit structure of its group members, a group's composition also influences non-group members. As mentioned before, the larger the number of contributors in an agent's group, the greater the agent's ability to withstand the pressure to change behavior. This makes empty spots near contributors highly valuable. The number of such spots changes substantially with the composition of existing groups (Figure 5), influencing the ability of non-contributors to freeride and the contributor's and non-contributor's ability to withstand pressure. This highlights an important tradeoff that exists at lower levels of contributors between the ability of an average agent to withstand pressure and that of its contributors which have managed to group with other contributors.

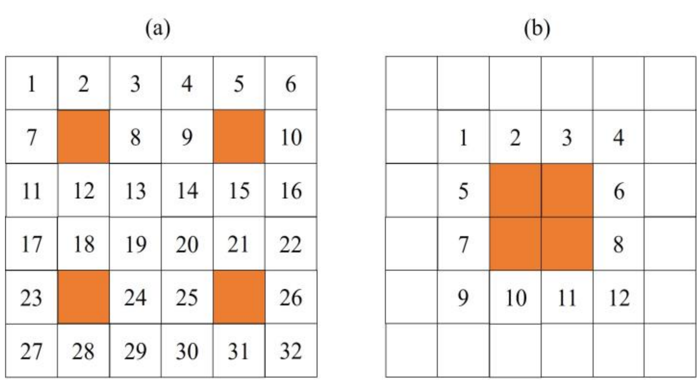

Figure 5: Two alternative group composition scenarios. Scenario (a) includes four spread out contributors (orange squares), offering 32 empty spots (8 each) next to them for other agents to fill. Scenario (b) includes four contributors clustered together - forming a square - offering only 12 empty spots next to them for others to fill. Scenario (a) benefits other agents, both contributors and non-contributors. Scenario (b) benefits the four contributors.

4.2 Macro-level analysis of model dynamics

4.2.1 Existence and classification of equilibria

The model is at equilibrium when all agents are satisfied with their group affiliation and behavior choice. If we assume homogeneous effort across all agents (prosocial or not), then both Equations 1 and 2 are linear in effort and without loss of generality this coefficient can be scaled to one. The value of prosocial common-pool behavior is then a linear function of the synergy and the proportion of the group adopting prosocial behavior. A freeriding group member receives this value plus the value of their own effort. In groups with no prosocial agents, the focal member only receives the value of their own effort, which is identical to the case of a lone agent without a group. The value of the prosocial common-pool behavior is therefore the entire potential benefit of group membership.

For the model to be at equilibrium, either there are no prosocial agents and the pressure is less than or equal to the individual effort (extinction), or there exist agents with prosocial behavior and across all groups the minimum value of the prosocial common-pool behavior is greater than or equal to the pressure (persistence or full adoption). Based on these observations, we can characterize the existence of different equilibria across the synergy-pressure parameter space (Figure 6).

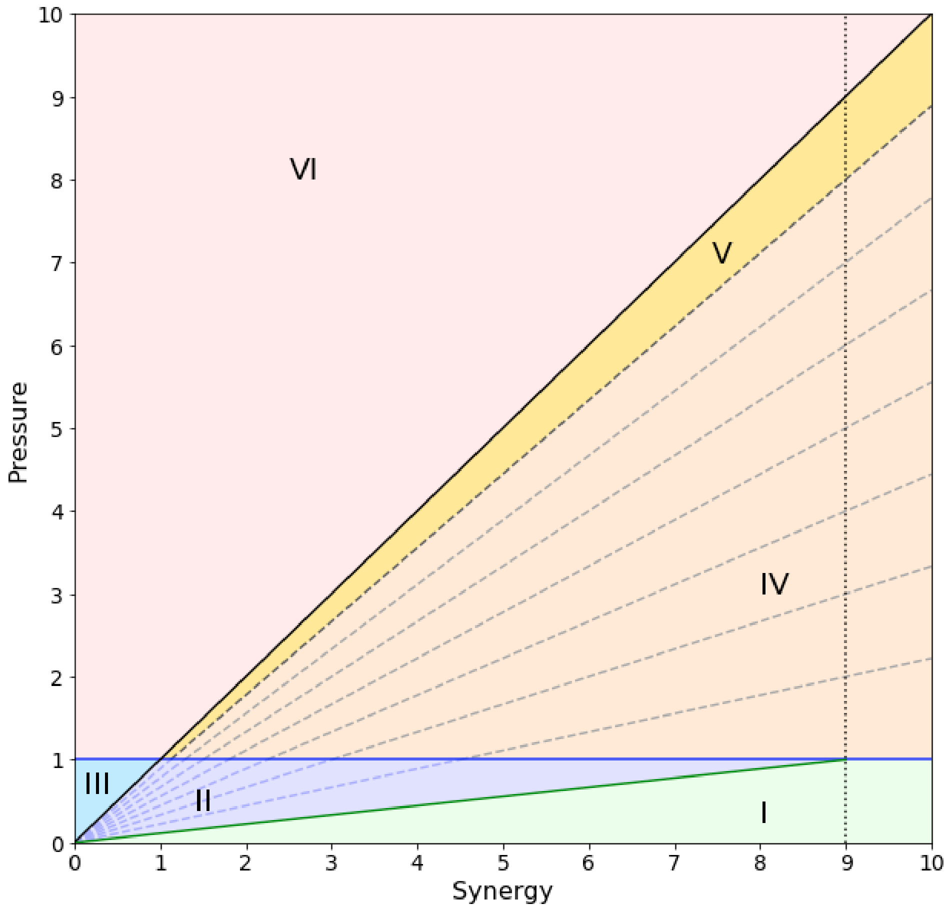

Figure 6: Parameter space map of model equilibria by region (**I-VI**). The figure was generated with Python's matplotlib package (version 3.4.1; Hunter, 2007).

Key elements in Figure 6 include: (a) the satisfaction boundary for lone agents and groups without prosocial behavior, $p = 1$; (b) the satisfaction boundary for contributing agents in groups with 100% prosocial behavior, $p = s$; and (c) the synergy level above which all agents will be better off adopting prosocial common-pool behavior regardless of the composition of their group, $s = 9$. We can break up the possible equilibria into six general regions.

Regions **I** through **III** are all below the satisfaction boundary for non-contributing agents, so no non-contributing agent will move or switch behavior in these regions. In region **I,** all behavior choices and group compositions produce sufficient reward to keep agents from moving or changing behavior. Pressure is too low to produce dynamics. All initial agent distributions are at equilibrium (persistence of prosocial common-pool behavior).

Region **II** has enough pressure to cause contributing agents without sufficient prosocial participation in their group to move and change behavior. Non-contributing agents are always satisfied whether or not their groups provide them benefit as freeriders. Dynamics from typical random initial distributions will exhibit short transients during which contributing agents move and change behavior. Most observed equilibria contain few contributing agents in stable groups (persistence of prosocial common-pool behavior). The required proportion of contributing agents for a stable group increases with pressure.

Region **III** is above and to the left of the $p = s$ limit for contributing agents to be at equilibrium, so every contributing agent will move and change behavior. All initial distributions lead to 100% non-contributing behavior after one iteration (extinction of prosocial common-pool behavior).

Regions **IV** through **VI** are all above the satisfaction boundary for non-contributing agents, which implies they cannot withstand the pressure to change without a freerider benefit in these regions. As region **VI** is also above and to the left of the $p = s$ limit for contributing agents to be at equilibrium, no equilibria exist in this region (persistence of prosocial common-pool behavior).

Region **IV** admits the existence of stable groups with contributing agents and freeriding non-contributing agents. The degree of freeriding permissible without destabilizing a group depends on the pressure. Dynamics from typical random initial distributions may exhibit short or long transients before reaching an equilibrium or the model dynamics may lead to the formation of stable groups in a configuration where one or more agents are trapped in a confined social space with no accessible satisfactory behavior and group combination available. In region **V** only 100% prosocial groups are stable (full adoption of common-pool behavior). There is no opportunity for freeriding at equilibrium.

4.2.2 Characterization of dynamics in absence of equilibria

Model dynamics in region **VI** fall in to two distinct categories, dynamics in which every agent moves and switches behavior on every iteration due to pressure exceeding the maximum common-pool benefit a freerider can receive (Region **VI A** in Figure 7), and dynamics in which freeriding agents can sustain their behavior for an iteration due to participation in a sufficiently prosocial group (Region **VI B** in Figure 7). The sufficiently prosocial groups that permit freeriders in region

VI B cannot persist as their members with prosocial behavior are unable to resist the pressure to move and change behavior.

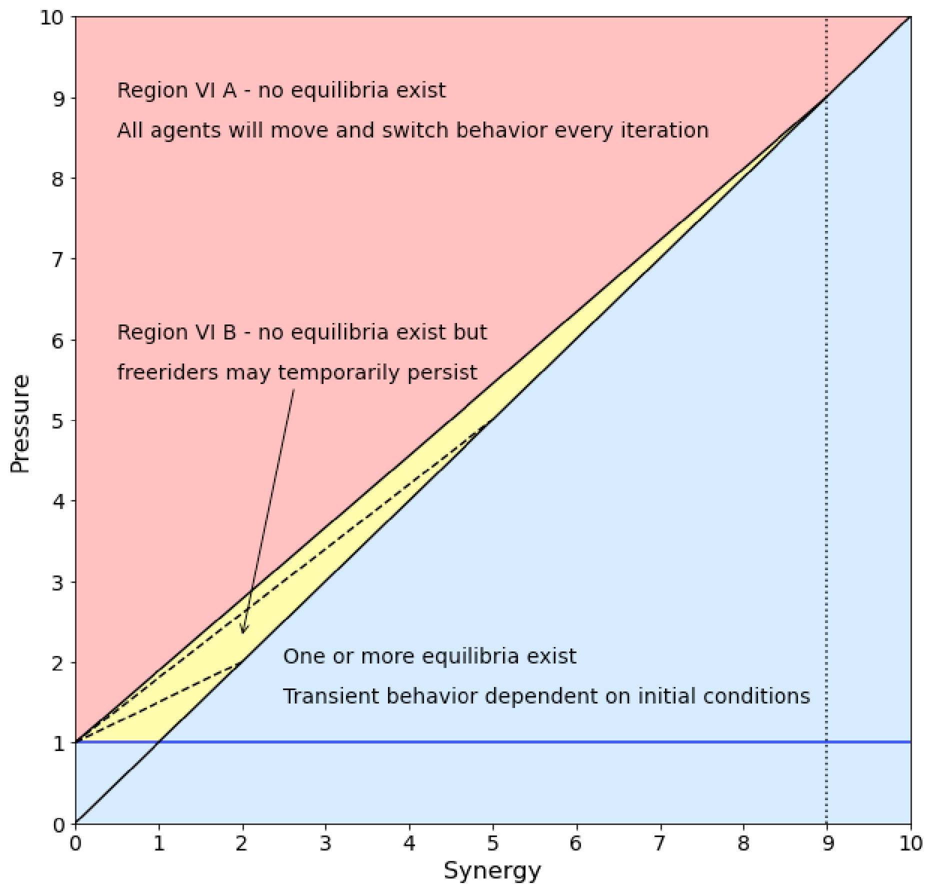

Figure 7: Parameter space map of model dynamics. The figure was generated with Python's matplotlib package (version 3.4.1; Hunter, 2007).

The ultimate boundary for region **VI B** is given by the maximum benefit for a freerider, $1 + (8/9) * s$. Obtaining this payoff requires a maximal sized group of contributing agents. Smaller groups may temporarily support freeriders when the pressure is closer to the region **VI** lower boundary as illustrated by the dashed lines in region **VI B**, which represent the payoff boundaries for groups with 80% and 50% contributing agents.

5 Simulation results

MGS I's structure allows for the following three alternative and substantially different outcomes:

1. When one or more equilibria are present, dynamics from typical random initial distributions may exhibit short or long transients before reaching an equilibrium (regions **I-V**).
2. In some regions, even though one or more equilibria exist, transients may lead to the formation of stable groups in a configuration where one or more agents are trapped in a confined social space with no accessible satisfactory behavior and group combination available, which leads to a persistent random cycle through a subregion of state space containing no equilibria (region **IV**).
3. When the pressure is high enough that no equilibria exist (region **VI**), all agents change their behavior either every iteration period, or some freeriders persist for one or more iterations before switching behaviors in a perpetual random walk.

5.1 Sweeping through the parameter space

To verify the theoretical analysis in Section 4 and further explore the relationship between simulation outcomes and specific parameter-setting combinations, we developed a variant of MGS I, called MGS I Sweep (version 1.0), which sweeps through a range of pressure-synergy combinations and, to account for stochasticity, does each sweep a preset number of times (which we set to 3) and then averages the results. We used MGS I Sweep to sweep through pressure-synergy combinations ranging from 0-0 to 10-10, at 0.1 increments, and at two alternative density (0.3, 0.7) and movement (move, no move) settings (Figure 8). We implemented the two movement settings by deactivating movement and re-sweeping through the same ranges of pressure-synergy combinations.

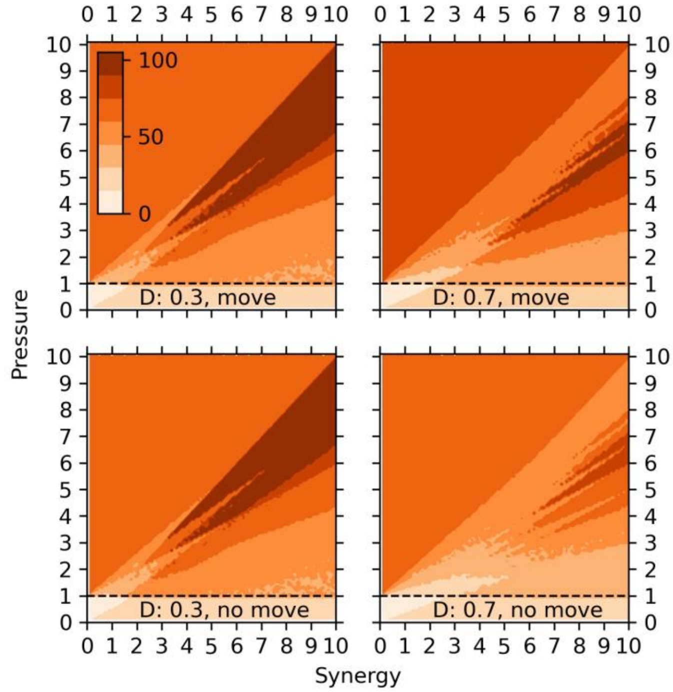

Figure 8: Four contour maps of MGS I's parameter space of percent of contributions after 1,000 iterations. Each map illustrates the results (final percent of contributors) of 10,000 model runs under alternative pressure-synergy combinations (each map is actually based on 30,000 model runs, since each combination was simulated three times and averaged to account for stochasticity). The maps differ in density (D: 0.3, 0.7) and movement (move, no move) settings, with the following parameter settings kept constant: $Z^2 = 400$, e = 1. The black horizontal dashed line depicts where effort equals pressure ($e = p = 1$). The figure was generated with Python's matplotlib package (version 3.4.1; Hunter, 2007).

The dynamics that lead to the outcomes that are visualized by the contour maps (Figures 8) are explained by the regions described in Section 4.2. The areas where equilibria are reliably reached include region **III** (extinction of prosocial common-pool behavior), which is clearly visible near the origin. Regions **I** and **II** also reach equilibria (persistence of prosocial common-pool behavior) but are not well differentiated by looking at the final percent of contributors due to the percent of contributors in the initial state. Some loss of contributors from the initial state is visible in that part of region **II** closest to the $s = p$ boundary.

Regions **IV** and **V** show where model dynamics lead to population-wide adoption of prosocial common-pool behavior and where the dynamics lead to partial adoption (Case 1 above) or partial adoption with some agents trapped in untenable social spaces (Case 2). Note that 100% prosocial behavior is always an equilibrium state in regions **IV** and **V**. The contours in regions **IV** and **V** can be described using the stability boundaries of various sized prosocial groups and freerider payoffs, and represent both types of dynamics.

Finally, the third case of no equilibrium is easily spotted as the upper left regions in all 4 maps corresponding to region **VI A** of our analysis. Region **VI B** is noticeable on both low-density maps (left top and bottom) as the area between region **VI A** and the $s = p$ boundary, and though less obvious, is also present in the high density maps.

5.2 The influence of population density

The influence of population density on social life is multidimensional and complex (e.g., Ahlfeldt & Pietrostefani, 2019), and can be simultaneously positive (increased innovation, reduced per capita pollution, access to amenities) and negative (higher congestion, and prices of land and housing) (Duranton & Puga, 2020). Its influence on prosocial common-pool behavior, however, is not widely (if at all) studied. In the model, population density influences the conditions that agents face during a simulation in two important ways. One is it influences the number of agents being simulated. This is because the number of agents is determined by the product of density and the number of spots in the social space ($D*Z^2$). An increase in density increases the chance of having a neighbor, which, from the perspective of an agent, has both positive and negative side effects. The positive side effect is the corresponding increase in the chance of having a contributor as a neighbor. The negative side effect is the increase in the chance of a neighbor being a non-contributor. The second way in which density influences the conditions that agents face is through the number of possible moves during a simulation (which density reduces) and thereby the chance of finding an empty spot near a contributor or non-contributor (for better or for worse, respectively).

The influence of population density on population-wide adoption of prosocial common-pool behavior is illustrated by the differences in the contour maps that were generated under the two alternative density settings (0.3, 0.7) (Figure 8). Overall, increases in density reduce the number of pressure-synergy combinations that lead to population-wide adoption of prosocial common-pool behavior. One likely cause is that higher population density levels increase the chance of having a non-contributing (i.e., a freeriding) neighbor, which reduces the ability of their neighbors to withstand pressure to change. Because within any group a contributor is at a disadvantage to a non-contributor, this effect has a greater negative impact on the former than the latter.

5.3 The influence of the initial percent of contributors

The relative numbers of socially or culturally different individuals in a population are critical in shaping their interaction and, in turn, future outcomes (Kanter, 1977). In the model, the influence of the initial percent of contributors on population-wide adoption of prosocial common-pool behavior affects underlying behavior and the final outcomes, but not the size of the regions or the transition zones between them (as the case with density). Final outcomes remain the same in regions **III** and **V**, in which the final percent of contributors reach 0% and 100%, respectively when equilibrium is reached. In region **I**, the final percent of contributors is simply equal to the initial percent as all initial conditions are at equilibrium. Equilibria obtained for region **II** are dependent on initial conditions as the initial percent is an upper limit on the final percent.

In region **VI A**, the initial percent of contributors influences the position of the two equilibria points between which the percent of contributors fluctuates during a simulation. Specifically, the locations of the equilibria points are symmetrically distanced from 50%. For example, if the initial percent of contributors is 40% (or 60%), the percent of contributors during a simulation fluctuates between 40% and 60% (or vice versa). If the initial percent is 30% (or 70%), the percent fluctuates between 30% and 70% (or vice versa). In all cases, the average final percent of contributors is 50%. However, depending on the time period at which a simulation ends, the final percent of contributors can be equal to either one of the equilibria. For example, the simulations that Figure 8 relies on ran for 1,000 time periods. With the simulations starting on an odd number (1) and ending on an even number (1,000), the final percent of contributors are equal to the other equilibrium point, e.g., if the initial percent of contributors is 30%, the final percent of contributors is 70%, and vice versa.

When interpreting these identical and symmetric fluctuations in population-wide behavior, it is important to keep in mind that we are describing the base model, based on which more realistic variants can be built. For example, a variant that assigns a probability to whether an agent changes its behavior would remove the identical and symmetric nature of the described fluctuations in the percent of contributors. Aside from their identical and symmetric nature, however, population-wide shifts in behavior as demonstrated by MGS I are anything but absent from human history. Primary examples include cultural and political revolutions (Kuran, 1995; Opp & Gern, 1993), which are increasingly “scaled up” by social media (Mundt et al., 2018). The aforementioned equilibria can be seen as “tipping points” (Centola et al., 2018) that lead to major shifts in social convention.

5.4 The influence of movement

Movement in MGS I represents any process of leaving and/or joining a group. Agents are selected randomly one-by-one and automatically move to one of the closest empty spots if their ability to withstand pressure is below the pressure level. Even when random, movement assists adoption of prosocial common-pool behavior. This is evident by the *all-around* lower final percent of contributors in the bottom row of Figure 8 contour plots, which are based on simulations without movement, when compared to those generated with the same density levels and initial percent of contributors, but with movement in the top row. This result challenges the recently proposed notion that greedy (i.e., goal-oriented, non-random) movement is necessary for increases in adoption of prosocial common-pool behavior (e.g., Helbing et al., 2011).

6 Discussion

6.1 Persistence and adoption of prosocial common-pool behavior

Individuals providing prosocial common-pool behavior are often at a substantial disadvantage to those who exploit it. Yet, prosocial common-pool behavior persists in most populations and new examples appear and are often adopted. MGS I offers a plausible explanation for why this occurs. It demonstrates how at sufficiently high levels of synergy from prosocial common-pool behavior, groups with larger numbers of contributors are better able to withstand pressure than groups with lower numbers. This mechanism is commonly-referred to in biology as multilevel group selection (Boyd, 2018; Goodnight & Stevens, 1997; Sober & Wilson, 1999; Waring et al., 2015; Wilson & Sober, 1994; Wilson & Wilson, 2007). Outside of a few exceptions (e.g., Boyd, 2018; Waring et al., 2015), models developed for studying multilevel group selection are focused on non-human populations. Therefore, one of our main contributions is a new base model for studying multilevel group selection in human populations.

The base model establishes a previously unexplored relationship among important factors influencing prosocial common-pool behavior, including group composition, group size, movement between groups, multilevel group selection, population density, pressure to change behavior, and the synergy from prosocial common-pool behavior. Our theoretical analysis of agent behavior (Section 4.1) described the influence: (a) size and synergy have on the types of situations contributors and non-contributors of prosocial common-pool behavior might face, (b) composition and synergy have on the ability of the contributors and non-contributors to withstand pressure, and (c) composition and pressure have on the benefit of joining a group.

Our theoretical analysis of the model's dynamics (Section 4.2) identified six distinct regions in the model's parameter space, in which pressure-synergy combinations lead to substantially different outcomes: extinction (region **III**), persistence (regions **I** & **II**), and full adoption (regions **IV** & **V**) of a prosocial common-pool behavior. Regions **IV** & **V** demonstrate the presence in the model of multilevel group selection: under sufficiently high synergy from prosocial common-pool behavior, adopters organize into sufficiently large groups that permit them to offset the disadvantage they have against non-adopters in withstanding pressure to change their behavior and lead to its population-wide adoption.

Our analysis of the model's simulation results (Section 5) verified the theoretical analysis in Section 4.2 and provided additional insight into the influence of the aforementioned factors on population-wide adoption of prosocial common-pool behavior. It demonstrated that: (a) increases in density reduce the number of pressure-synergy combinations that lead to population-wide adoption of prosocial common-pool behavior, (b) the initial percent of contributors affects the underlying behavior and the final outcomes, but not the size of the regions or the transition zones between them, and (c) random (non-greedy) movement assists adoption of prosocial common-pool behavior. Overall, the analyses demonstrate that further exploration of the relationships among the important factors has the potential to be of great value in efforts to promote prosocial common-pool behavior.

6.2 Relevance for managing tangible common-pool resources

The MGS I model also has the potential to provide insight into the management of common-pool resources (e.g., fisheries, forests, groundwater basins). The benefit derived from consuming a common-pool resource is private to the consumer, while the cost associated with its consumption is shared by all consumers (Hardin, 1968). For this reason, management efforts aimed at preventing exploitation of a tangible common-pool resource tend to fall into one or a combination of the following two categories: (a) efforts that limit an individual's consumption and (b) those that force an individual to internalize the associated cost.

However, outside of a perfect market economy, the decision to consume a tangible common-pool resource is rarely ever solely based on the aforementioned cost-benefit structure alone. Resisting overconsuming is a prosocial common-pool behavior that, as we have demonstrated in Section 4.1, is encouraged/discouraged by: (a) the broader synergy from the prosocial common-pool behavior to the individual and (b) the pressure to change their behavior. Both include the benefit and cost associated directly with the consumption of a resource. As mentioned in Section 2, prosocial common-pool behavior might additionally and, perhaps, in some cases, be entirely motivated by altruism, egoism, or by a desire to adhere to a personal ideology. As demonstrated in Section 5, social structure, movement, and multilevel group selection also influence the prevalence of prosocial common-pool behavior.

Therefore, populations aiming to prevent exploitation of tangible common-pool resources have additional psychological and social levers that complement those related to limiting consumption of the resource and internalizing its cost. Many social levers are already incorporated in formal and informal norms. Examples include social punishment for overconsumption and social reward for restraint. Many psychological levers are also already ingrained during upbringing (e.g., Simon & Mobekk, 2019) and in some cases are also genetically transmitted through prosocial genes (Knafo-Noam et al., 2015). The analyses in Sections 4.2 and 5 suggest that the right pressure-synergy combinations can lead populations to higher levels of adoption of prosocial common-pool behavior, and, through association, responsible consumption of tangible common-pool resources. It additionally suggests that social structure plays a considerable role in encouraging/discouraging adoption.

6.3 Potential variants

MGS I has a simple and modular structure that eases the process of developing its variants that replace existing processes, add complexity to them, and/or introduce new processes while maintaining important components in situations with prosocial common-pool behavior, i.e., density, movement, multilevel group selection, pressure, and synergy. We experimented with a number of what would now be considered as variants in the process of designing MGS I, including ones that introduce: greedy (i.e., goal-oriented, non-random) movement, probability in behavioral change, and limitations of group access to non-contributors. Each one of these added a layer of real-world complexity that influenced model results, but were excluded from MGS I in pursuit of a fundamental design.

Here, we briefly mention two possible directions for future variants not mentioned above. The first direction involves variants that empower MGS I agents with an ability to assess the likely behavior of group members and choose their own group behavior accordingly. These variants would permit the study of game theoretic situations in a population consisting of multiple groups of varying size. The new processes would replace the existing process of behavioral change, which automatically changes an agent's behavior when pressure is equal to or above its ability to withstand it. To study the Prisoner's Dilemma or the coordination game, specifically, a variant would additionally need to limit the ranges of the pressure and synergy parameters, as well as introduce a mechanism for regulating group size (see Section 4.1.1). A resulting variant could then be used to explore games in multi-group populations under alternative pressure-synergy combinations, and density and initial percent of contributors settings.

The second direction involves variants that empower MGS I agents with a wider range of cognitive, social, demographic, and potentially additional behavioral processes. These variants would permit the study of the influence of these processes on adoption of prosocial common-pool behavior. One option could be to replace the base model's agents with SOSIEL agents (Sotnik, 2018), each of which makes decisions using a cognitive architecture composed of nine processes (anticipatory learning, goal prioritizing, counterfactual thinking, innovating, social learning, goal selecting, satisficing, signaling, and individual or collective action taking) and can be born, die, and pair with other agents. This second direction could of course be configured to also align with the first direction, permitting the study of the influence of cognitive, behavioral, social, and demographic processes on adoption of prosocial common-pool behavior in game theoretic situations. These aforementioned directions for future variant development are only two among many, serving as examples of how MGS I provides a fertile base for a fruitful and socially-relevant research program.

References

Abrahamson, E. (1991). Managerial Fads and Fashions: The Diffusion and Rejection of Innovations. *The Academy of Management Review*, *16*(3), 586–612.

Ahlfeldt, G. M., & Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, *111*, 93–107. https://doi.org/10.1016/j.jue.2019.04.006

Axelrod, R. M. (1984). *The Evolution of Cooperation*. Basic Books.

Axelrod, R. M., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Science, New Series*, *211*(4489), 1390–1396.

Barnett, J., & O’Neill, S. (2010). Maladaptation. *Global Environmental Change*, *20*(2), 211–213. https://doi.org/10.1016/j.gloenvcha.2009.11.004

Battaglini, M., Nunnari, S., & Palfrey, T. (2012). *The Free Rider Problem: A Dynamic Analysis* (Working Paper No. 17926; Issue 17926). National Bureau of Economic Research. http://www.nber.org/papers/w17926

Bowles, S., & Gintis, H. (2009). Beyond Enlightened Self-Interest: Social Norms, Other-Regarding Preferences, and Cooperative Behavior. In S. A. Levin (Ed.), *Games, Groups, and the Global Good* (pp. 57–78). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85436-4

Boyd, R. (2018). *A different kind of animal: How culture transformed our species*. Princeton University Press.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*(3), 171–195.

https://doi.org/10.1016/0162-3095(92)90032-Y

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Centola, D., Becker, J., Brackbill, D., & Baronchelli, A. (2018). Experimental evidence for tipping points in social convention. *Science*, *360*(6393), 1116–1119. https://doi.org/10.1126/science.aas8827

Choi, T., & Robertson, P. J. (2019). Contributors and Free-Riders in Collaborative Governance: A Computational Exploration of Social Motivation and Its Effects. *Journal of Public Administration Research and Theory*, *29*(3), 394–413. https://doi.org/10.1093/jopart/muy068

Cornes, R., & Sandler, T. (1996). *The Theory of Externalities, Public Goods, and Club Goods*. Cambridge University Press.

Dal Bó, P., & Fréchette, G. R. (2019). Strategy Choice in the Infinitely Repeated Prisoner’s Dilemma. *American Economic Review*, *109*(11), 3929–3952. https://doi.org/10.1257/aer.20181480

De Silva, H., & Sigmund, K. (2009). Public Good Games with Incentives: The Role of Reputation. In S. A. Levin (Ed.), *Games, Groups, and the Global Good* (pp. 85–103). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85436-4

Duranton, G., & Puga, D. (2020). The Economics of Urban Density. *Journal of Economic Perspectives*, *34*(3), 3–26. https://doi.org/10.1257/jep.34.3.3

Frank, S. A. (2009). Evolutionary Foundations of Cooperation and Group Cohesion. In S. A. Levin (Ed.), *Games, Groups, and the Global Good* (pp. 3–40). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-85436-4

Gilbert, P., & Basran, J. (2019). The Evolution of Prosocial and Antisocial Competitive Behavior and the Emergence of Prosocial and Antisocial Leadership Styles. *Frontiers in Psychology*, *10*, 610. https://doi.org/10.3389/fpsyg.2019.00610

Goodnight, C. J., & Stevens, L. (1997). Experimental Studies of Group Selection: What Do They Tell US About Group Selection in Nature? *The American Naturalist*, *150*(S1), S59–S79. https://doi.org/10.1086/286050

Hardin, G. (1968). The Tragedy of the Commons. *Science*, *162*(3859), 1243–1248. https://doi.org/10.1126/science.162.3859.1243

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K. (2002). Replicator Dynamics for Optional Public Good Games. *Journal of Theoretical Biology*, *218*(2), 187–194. https://doi.org/10.1006/jtbi.2002.3067

Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A., & Sigmund, K. (2007). Via Freedom to Coercion: The Emergence of Costly Punishment. *Science*, *316*(5833), 1905–1907. https://doi.org/10.1126/science.1141588

Helbing, D., & Yu, W. (2008). Migration as a mechanism to promote cooperation. *Advances in Complex Systems*, *11*(04), 641–652. https://doi.org/10.1142/S0219525908001866

Helbing, D., Yu, W., & Rauhut, H. (2011). Self-Organization and Emergence in Social Systems: Modeling the Coevolution of Social Environments and Cooperative Behavior. *The*

Journal of Mathematical Sociology, *35*(1–3), 177–208. https://doi.org/10.1080/0022250X.2010.532258

Hess, C., & Ostrom, E. (2003). Ideas, Artifacts, and Facilities: Information as a Common-Pool Resource. *Law and Contemporary Problems*, *66*(1/2), 111–145.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Juhola, S., Glaas, E., Linnér, B.-O., & Neset, T.-S. (2016). Redefining maladaptation. *Environmental Science & Policy*, *55*, 135–140. https://doi.org/10.1016/j.envsci.2015.09.014

Kanter, R. M. (1977). Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women. *American Journal of Sociology*, *82*(5), 965–990. https://doi.org/10.1086/226425

Kim, O., & Walker, M. (1984). The free rider problem: Experimental evidence. *Public Choice, 43*(1), 3–24. https://doi.org/10.1007/BF00137902

Knafo-Noam, A., Uzefovsky, F., Israel, S., Davidov, M., & Zahn-Waxler, C. (2015). The prosocial personality and its facets: Genetic and environmental architecture of mother-reported behavior of 7-year-old twins. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00112

Kuran, T. (1995). The Inevitability of Future Revolutionary Surprises. *The American Journal of Sociology*, *100*(6), 1528–1551.

Magnan, A. K., Schipper, E. L. F., Burkett, M., Bharwani, S., Burton, I., Eriksen, S., Gemenne, F., Schaar, J., & Ziervogel, G. (2016). Addressing the risk of maladaptation to climate change: Addressing the risk of maladaptation to climate change. *Wiley Interdisciplinary*

Reviews: Climate Change, *7*(5), 646–665. https://doi.org/10.1002/wcc.409

Maskin, E. (2016). How Can Cooperative Game Theory Be Made More Relevant to Economics? : An Open Problem. In J. F. Nash, & M. Th. Rassias (Eds.), *Open Problems in Mathematics* (pp. 347–350). Springer International Publishing. https://doi.org/10.1007/978-3-319-32162-2_9

Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling Social Movements Through Social Media: The Case of Black Lives Matter. *Social Media + Society*, *4*(4), 205630511880791. https://doi.org/10.1177/2056305118807911

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, *359*.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, *393*, 573–577.

Opp, K.-D., & Gern, C. (1993). Dissident Groups, Personal Networks, and Spontaneous Cooperation: The East German Revolution of 1989. *American Sociological Review*, *58*(5), 659. https://doi.org/10.2307/2096280

Ostrom, E. (1999). Revisiting the Commons: Local Lessons, Global Challenges. *Science*, *284*(5412), 278–282. https://doi.org/10.1126/science.284.5412.278

Ostrom, E., Gardner, R., & Walker, J. (Eds.). (1994). *Rules, Games, and Common-Pool Resources*. The University of Michigan Press.

Ozono, H., Kamijo, Y., & Shimizu, K. (2017). Punishing second-order free riders before first-order free riders: The effect of pool punishment priority on cooperation. *Scientific Reports*, *7*(1), 14379. https://doi.org/10.1038/s41598-017-13918-8

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed). Free Press ; Collier Macmillan.

Simon, C., & Mobekk, H. (2019). Dugnad: A Fact and a Narrative of Norwegian Prosocial

Behavior. *Perspectives on Behavior Science*, *42*(4), 815–834. https://doi.org/10.1007/s40614-019-00227-w

Sober, E., & Wilson, D. S. (1999). *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press.

Sotnik, G. (2018). The SOSIEL Platform: Knowledge-based, cognitive, and multi-agent. *Biologically Inspired Cognitive Architectures*, *26*, 103–117. https://doi.org/10.1016/j.bica.2018.09.001

Tarnita, C. E., Antal, T., Ohtsuki, H., & Nowak, M. A. (2009). Evolutionary dynamics in set structured populations. *Proceedings of the National Academy of Sciences*, *106*(21), 8601–8604. https://doi.org/10.1073/pnas.0903019106

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., & Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *Journal of Personality and Social Psychology*, *92*(1), 56–66. https://doi.org/10.1037/0022-3514.92.1.56

van Rossum, G., & Drake Jr., F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Wang, J., Wu, B., W. C. Ho, D., & Wang, L. (2011). Evolution of cooperation in multilevel public goods games with community structures. *EPL (Europhysics Letters)*, *93*(5), 58001. https://doi.org/10.1209/0295-5075/93/58001

Waring, T. M., Kline, M. A., Brooks, J. S., Goff, S. H., Gowdy, J., Janssen, M. A., Smaldino, P. E., & Jacquet, J. (2015). A multilevel evolutionary framework for sustainability analysis. *Ecology and Society*, *20*(2), art34. https://doi.org/10.5751/ES-07634-200234

Wilensky, U. (1999). *NetLogo* (6.2.0) [NetLogo]. Center for Connected Learning and Computer-

Based Modeling, Northwestern University. http://ccl.northwestern.edu/netlogo/

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, *17*(4), 585–608. https://doi.org/10.1017/S0140525X00036104

Wilson, D. S., & Wilson, E. O. (2007). Rethinking the Theoretical Foundation of Sociobiology. *The Quarterly Review of Biology*, *82*(4), 327–348. https://doi.org/10.1086/522809

Ye, H., Tan, F., Ding, M., Jia, Y., & Chen, Y. (2011). Sympathy and Punishment: Evolution of Cooperation in Public Goods Game. *Journal of Artificial Societies and Social Simulation*, *14*(4), 20. https://doi.org/10.18564/jasss.1805