



# Green-beard effect predicts the evolution of traitorousness in the two-tag Prisoner's dilemma

Robert A. Laird\*

Department of Biological Sciences, University of Lethbridge, Lethbridge, Canada AB T1K 3M4

## ARTICLE INFO

### Article history:

Received 18 February 2011  
Received in revised form  
12 July 2011  
Accepted 25 July 2011  
Available online 2 August 2011

### Keywords:

Game theory  
Evolutionary dynamics  
Extra-tag cooperation  
Prisoner's dilemma  
Tag-based cooperation

## ABSTRACT

Cooperation, a costly interaction in which individuals benefit one another, plays a crucial role in many of the major transitions of evolution. Yet, as illustrated by the Prisoner's dilemma, cooperative systems are fragile because cooperators can be exploited by defectors who reap the benefits of cooperation but do not reciprocate. This barrier to cooperation may be overcome if cooperators have a recognisable phenotypic tag that allows them to adopt the conditional strategy of cooperating with fellow tag-mates while defecting against others, a mechanism known as the 'green-beard effect'. The resulting intra-tag cooperator strategy is particularly effective in structured populations where local clumps of cooperative tag-mates can find refuge. While intra-tag cooperation is robust against unconditional defectors in the spatial Prisoner's dilemma (at least when the cost of cooperation is low), the role of extra-tag cooperators – individuals who cooperate only with those bearing a different tag – has received little attention, despite the fact that these traitors form mixed-tag aggregations whose heterogeneous makeup potentially allows the exploitation of multiple other strategies. Using a spatial model of the two-tag Prisoner's dilemma, I show that extra-tag cooperation readily evolves under low to intermediate cost–benefit ratios of mutual cooperation ( $r$ ). Specifically, at low  $r$ , mixed-tag aggregations of extra-tag cooperators take over the population, while at intermediate  $r$ , such aggregations coexist with intra-tag cooperators and unconditional defectors with whom they engage in non-transitive spatial invasibility. In systems with more than two tags, however, the dilution of extra-tag cooperators within mixed-tag aggregations prevents the strategy from being effective. Thus, the same beard chromodynamics that promotes within-group cooperation also predicts the evolution of traitorous between-group cooperation, but only when the number of beard colours is low.

© 2011 Elsevier Ltd. All rights reserved.

## 1. Introduction

Cooperation, a behaviour in which an individual provides a benefit to another at a cost to itself, is instrumental to many of the major transitions in evolution, including those from genes to genomes, from single- to multi-cellular organisms, and from individuals to societies (Maynard Smith and Szathmáry, 1997; Nowak, 2006a). Yet cooperation is a long-standing puzzle in evolutionary biology because cooperators are vulnerable to exploitation by defectors—cheaters who accept the benefits of cooperation from others but fail to reciprocate. The paradigmatic Prisoner's dilemma game distils this puzzle to its essence and is the main theoretical framework used to study the evolutionary game dynamics of cooperation (Nowak, 2006b). As with fixed-strategy, two-player games in general, players in the Prisoner's dilemma either cooperate with or defect against their co-player. The game corresponds to a

situation where cooperation is beneficial yet costly, and therefore easily undermined by opportunistic defectors. Formally, cooperators provide a benefit  $b$  to their co-player at a cost of  $c$  to themselves ( $b > c > 0$ ). Defectors provide no benefit and pay no cost. The game's net payoffs, which are translated into evolutionary fitness, depend on the combination of strategies the players adopt (Maynard Smith, 1982). In the case of mutual cooperation, both players receive the benefit and pay the cost of cooperation, and therefore receive a payoff of  $R = b - c$ . Mutual defection results in a payoff of  $P = 0$ . When one player cooperates and the other defects, the defector gets the benefit without paying the cost for a payoff of  $T = b$ ; the cooperator, on the other hand, pays the cost without getting the benefit for a payoff of  $S = -c$ . In large, well-mixed populations, an individual attains greater fitness by defecting than by cooperating, regardless of whether their co-player cooperates ( $T > R$ ) or defects ( $P > S$ ). Thus, defection is an evolutionarily stable strategy, and evolutionary dynamics predicts the demise of cooperation, despite the fact that a population of cooperators has a greater mean fitness than a population of defectors ( $R > P$ ) (Nowak, 2006b; Maynard Smith, 1982; Axelrod and Hamilton, 1981; Axelrod, 2006).

\* Tel.: +1 403 317 5074.

E-mail address: [robert.laird@uleth.ca](mailto:robert.laird@uleth.ca)

Attempts to reconcile this conclusion with observed widespread cooperation in nature (Sherratt and Wilkinson, 2009) have focused on mechanisms that allow cooperators to interact with one another assortatively (Nowak and Sigmund, 2004), i.e., more frequently than predicted by mean-field considerations. Thus, cooperation can evolve via iterated interactions (Axelrod and Hamilton, 1981), reputational effects (Nowak and Sigmund, 1998), interactions that occur on local networks (Nowak and May, 1992; Lieberman et al., 2005) (which can be viewed as a form of kin selection caused by limited dispersal (Taylor et al., 2007)), and group selection (Traulsen and Nowak, 2006) (reviewed in Nowak (2006a)). One particularly simple way of aligning cooperative behaviour is by recognition. The existence of arbitrarily different types of individuals allows for recognition, and opens up the possibility of the emergence of more complex strategies than simple cooperation and defection. For instance, discriminating individuals might cooperate with members of some definable group to which they themselves belong, while defecting against non-members. Indeed the ‘green beard’ hypothesis proposes that a gene simultaneously coding for (i) a heritable, identifiable ‘tag’ (the proverbial green beard), (ii) the ability to recognise the tag, and (iii) the propensity to help others bearing the tag (or harm those not bearing the tag (Lehmann et al., 2009)), could allow for the evolution of cooperation (Hamilton, 1964; Dawkins, 1976; Gardner and West, 2009; West and Gardner, 2010). Originally, this tag-based cooperation mechanism was deemed unlikely on the grounds that it seemed implausible that a single gene could exhibit such broadly pleiotropic effects (Hamilton, 1964; Dawkins, 1976) (but see Keller and Ross (1998); Queller et al. (2003); Smukalla et al. (2008), and other examples cited by Gardner and West (2009); West and Gardner (2010)). Furthermore, if the suite of traits were instead controlled by multiple genes, then a population of green-bearded cooperators would be subject to invasion by green-bearded defectors, i.e., those that bear the ostensible cooperation tag but fail to cooperate (Dawkins, 1982).

Nevertheless, a number of recent advances have placed tag-based cooperation on a stronger theoretical foundation (Gardner and West, 2009; Riolo et al., 2001; Roberts and Sherratt, 2002; Traulsen and Schuster, 2003; Axelrod et al., 2004; Hammond and Axelrod, 2006a, 2006b; Jansen and van Baalen, 2006; Masuda and Ohtsuki, 2007; Traulsen and Nowak, 2007; Traulsen, 2008; Antal et al., 2009; Shultz et al., 2009), with the key innovation being the allowance of multiple different tags (e.g., beard colours) to co-occur within the same population. This scenario admits four basic fixed-strategy types (Axelrod et al., 2004): unconditional cooperators (*C*) who cooperate with everyone indiscriminately, unconditional defectors (*D*) who defect against everyone, ‘nepotistic’ intra-tag cooperators (*I*) who cooperate only with tag-mates, and ‘traitorous’ extra-tag cooperators (*E*) who cooperate only with non-tag-mates. It is important to emphasise that *E* strategists are fundamentally different from ‘conditional harmers’ that have been examined elsewhere (Lehmann et al., 2009). Conditional harmers recognise and harm non-tag-mates, and are therefore nepotistic, rather than traitorous (i.e., they are like *I* strategists, except that with conditional harmers, benefits indirectly accrue to tag-mates by reducing competition from non-tag-mates).

Tags can be added to the notation as subscripts; e.g.,  $E_i$  is an extra-tag cooperator bearing Tag *i*. Thus, in interactions between individuals of the same tag, those using  $C_i$  or  $I_i$  strategies are situational cooperators, while those using  $E_i$  or  $D_i$  strategies are situational defectors. In contrast, in interactions between individuals of different tags, those using  $C_i$  or  $E_i$  strategies are situational cooperators, while those using  $I_i$  or  $D_i$  strategies are situational defectors. It follows that the payoff matrix for

tag-based interactions is given by

$$\mathbf{A} = \begin{matrix} & C_i & I_i & E_i & D_i & C_j & I_j & E_j & D_j \\ \begin{matrix} C_i \\ I_i \\ E_i \\ D_i \end{matrix} & \begin{pmatrix} R & R & S & S & R & S & R & S \\ R & R & S & S & T & P & T & P \\ T & T & P & P & R & S & R & S \\ T & T & P & P & T & P & T & P \end{pmatrix} & , & i \neq j \end{matrix} \quad (1)$$

where  $A_{pq}$  is the payoff to a focal individual with the strategy and tag corresponding to row *p* when playing against a competitor with the strategy and tag corresponding to column *q*. Given the definitions of Prisoner’s dilemma payoffs listed above, this four-parameter payoff matrix can be changed to a one-parameter matrix by substituting  $R=1$ ,  $P=0$ ,  $T=1+r$ , and  $S=-r$ , where  $r=c/(b-c)$  is the cost–benefit ratio of mutual cooperation (Doebeli and Hauert, 2005).

When only types *C* and *I* are present, green-beard cooperation readily evolves (Riolo et al., 2001; Traulsen and Schuster, 2003), assuming the two types bear a different tag. However, this may be undermined by the addition of *D* (Roberts and Sherratt, 2002) unless the population (i) occurs on sparsely connected regular or random networks (Hammond and Axelrod, 2006b; Jansen and van Baalen, 2006), (ii) experiences reputational effects (Masuda and Ohtsuki, 2007), (iii) is sufficiently small (in concert with a large *b/c* ratio) (Traulsen and Nowak, 2007; Traulsen, 2008), or (iv) experiences differential tag- and strategy-mutation rates (Antal et al., 2009). The role of *E* strategists in the evolution of cooperative behaviour has received far less attention than the roles of *C*, *I*, and *D* strategists (Axelrod et al., 2004; Hammond and Axelrod, 2006a, 2006b; Shultz et al., 2009). There are at least three reasons for this. The first is historical: *C*, *I*, and *D* have a close connection to the original green-beard theory, while *E* does not. Second, the *E* strategy, in its traitorousness, may appear to be an unviable proposition from the outset: in structured populations, *E* is merely as good as *D* when found within patches of tag-mates, yet unlike *D*, *E* is easily exploited at the margins of such patches by non-tag-mates (Shultz et al., 2009). However, this neglects the facts that (i) *E* strategists of different tags are mutually miscible, promoting the formation of mixed-tag aggregations whose heterogeneous composition potentially allows the exploitation of multiple other strategies (see Section 3.1), and, more generally, (ii) extending strategy space in evolutionary games often leads to counterintuitive results, such as the surprising emergence of anti-social punishment (Rand et al., 2010). Third, the few studies that allowed the possibility for extra-tag cooperation to evolve either did not report its evolution separately from other types of cooperators (Hammond and Axelrod, 2006b), or found it to be dominated by the other strategies (Axelrod et al., 2004; Hammond and Axelrod, 2006a; Shultz et al., 2009). Nevertheless, this purported dominance was typically based on one or two payoff matrices – i.e., a very limited range of the cost–benefit ratio, *r* – even though variation in *r* is critically important to the balance between strategy coexistence and exclusion in non-tag-based systems (Doebeli and Hauert, 2005; Hauert and Doebeli, 2004). Here, I report complementary spatial and aspatial agent-based models of the two-tag Prisoner’s dilemma that show that for spatially structured populations, the *E* strategy dominates *C*, *I*, and *D* – either quasi-stably or cyclically – provided that the cost–benefit ratio of mutual cooperation is sufficiently low. Interestingly, this mechanism for the evolution of extra-tag cooperation is ineffective when there are more than two tags present in the population, suggesting the hypothesis that green-beard cooperation may be relatively immune to traitorousness when a population’s ‘beard chromodiversity’ is high.

## 2. Methods

In the current models, which incorporate elements from previous work (Nowak and May, 1992; Hauert and Doebeli, 2004; Durrett and Levin, 1994), space is represented by a square  $100 \times 100$  lattice with periodic boundaries. A single individual belonging to a particular tag-by-strategy combination occupies each cell; there are no empty cells. Cells are selected randomly for potential replacement, with  $100^2$  such sequential events defined as one generation, so that every cell is selected for potential replacement once per generation, on average. When a focal cell is selected, a competitor cell is randomly selected from the four cells in the focal cell's von Neumann neighbourhood (Durrett and Levin, 1994). The average payoffs to focal and competitor cells from their respective von Neumann neighbourhoods are then compared; these are called  $p_x$  and  $p_y$ , respectively. If  $p_y \leq p_x$ , the occupant of the focal cell is unchanged. However, if  $p_y > p_x$ , an asexual clone of the occupant of the competitor cell replaces the occupant of the focal cell with probability  $(p_y - p_x)/(T - S)$  (Hauert and Doebeli, 2004). Finally, the focal cell mutates to a randomly selected tag and strategy with probability  $\mu$  (all transitions are equally likely). In most of the analyses presented here,  $\mu = 10^{-4}$ , corresponding to one mutant per generation, on average (other values of  $\mu$  are considered in the Appendices). The aspatial version of the model is identical, except all the cells (focal, competitor, and their respective neighbours) are chosen randomly from the whole population. The models were implemented in MATLAB.

## 3. Results and discussion

### 3.1. Pairwise strategy dynamics in spatially structured populations

In lattice models of spatial competition, invasion and replacement occur along borders between pairs of patches. Therefore, before considering populations in which all the tag-by-strategy types are potentially present, it is instructive to consider how pairs of types interact in the spatially explicit arena, in the absence of mutation. These situations are sufficiently simple that the outcome of competition can be deduced *a priori*; however, these deductions also match the results of simulations (Appendix A). Thus, Table 1a and b summarises the outcome of spatial

competition when only two tag-by-strategy types are present and there is no mutation:

Some pairs of tag-by-strategy types lead to spatial standoffs, where neither type can invade the other. This can occur when all interactions result in either mutual cooperation (e.g.,  $C_i$  versus  $I_i$ , or  $C_i$  versus  $C_j$ ) or mutual defection (e.g.,  $E_i$  versus  $D_i$ , or  $D_i$  versus  $D_j$ ), which means that every individual has the same payoff (Supplementary Figs. A.1, A.2, A.5, A.8). Another type of standoff occurs when  $C_i$  meets  $E_j$ . Initially  $C_i$  can outcompete  $E_j$ , but once the population reaches a configuration in which every  $E_j$  individual is surrounded by all  $C_i$  neighbours, every interaction results in mutual cooperation, leading to a standoff (Supplementary Figs. A.2, A.10). A weaker type of standoff can occur when  $I_i$  meets  $I_j$  (Supplementary Figs. A.2, A.11). Both  $I_i$  and  $I_j$  perform better with others of the same tag than with others of a different tag (respective payoffs  $R$  and  $P$ ). This results in large, monotypic, spatial aggregations of  $I_i$  and  $I_j$ . In the long term, this can lead to standoff or quasi-standoff spatial configurations (e.g., when there is a perfectly straight border between the two types). However, if by chance one of the types gains the upper hand and surrounds the other type, then this can lead to the latter's extinction, because in competition between different-tagged  $I$  strategists, individuals along a concave border have a greater average payoff than individuals along a convex border (Supplementary Figs. A.2, A.12). Similar 'standoff/replacement' dynamics have been noted elsewhere in other contexts (e.g., van Baalen and Jansen, 2003).

In other scenarios, one tag-by-strategy type always cooperates, while the other always defects (e.g.,  $C_i$  versus  $E_i$ ,  $C_i$  versus  $D_i$ ,  $I_i$  versus  $E_i$ ,  $I_i$  versus  $D_i$ , and  $C_i$  versus  $D_j$ ). Thus, these situations are exactly equivalent to a non-tag-based spatial Prisoner's dilemma with fixed strategies  $C$  and  $D$ . Provided  $r$  is sufficiently low (less than approximately 0.07, as demonstrated by Doebeli and Hauert (2005); Supplementary Figs. A.1, A.2, A.6), the situational cooperator can persist by forming compact clusters; otherwise the situational defector takes over (Supplementary Fig. A.7).

In still other cases, one tag-by-strategy type wins outright, regardless of the value of  $r$  (e.g.,  $I_i$  beats  $C_j$ ,  $E_j$ , and  $D_j$ ;  $D_i$  beats  $E_j$ ). This can occur in two ways. In the first way (e.g.,  $I_i$  beats  $C_j$  and  $E_j$ ;  $D_i$  beats  $E_j$ ), an individual of the winning tag-by-strategy type is able to exploit an individual of the losing type when they are adjacent (respective payoffs  $T$  and  $S$ ; Eq. (1)), and the winning type's average payoff with its other three neighbours is greater than or equal to the losing type's average payoff, regardless of the

**Table 1**  
(a,b) Spatial invasion dynamics when only two strategy-by-tag types are present and there is no mutation. (a) Outcome of spatial competition between members of the same tag-group and (b) outcome of spatial competition between members of different tag-groups. The 'blanks' along the diagonal in (a) denote the fact that these scenarios only have one type of individual and are therefore not relevant to the present analysis. Note that differently tagged  $E$  strategists are the only tag-by-strategy types capable of fine-grain mixing. (c) Outcome of spatial competition when there is a two-tag aggregation of  $E$  strategists ( $E_i E_j$ ), and one other tag-by-strategy type is present. The 'blank' in (c) denotes the fact that the column tag-by-strategy type is also a member of the multi-tag aggregation and is therefore subsumed by it. In all cases, the initial configuration was random.

(a) Same tag	$C_i$	$I_i$	$E_i$	$D_i$
$C_i$	–	Standoff	$E_i$ wins or coexistence	$D_i$ wins or coexistence
$I_i$	Standoff	–	$E_i$ wins or coexistence	$D_i$ wins or coexistence
$E_i$	$E_i$ wins or coexistence	$E_i$ wins or coexistence	–	standoff
$D_i$	$D_i$ wins or coexistence	$D_i$ wins or coexistence	Standoff	–
(b) Different tag	$C_j$	$I_j$	$E_j$	$D_j$
$C_i$	Standoff	$I_j$ wins	Isolation of $E_j$ , then standoff	$D_j$ wins or coexistence
$I_i$	$I_i$ wins	Standoff or extinction	$I_i$ wins	$I_i$ wins
$E_i$	Isolation of $E_i$ , then standoff	$I_j$ wins	Fine-grain mixing	$D_j$ wins
$D_i$	$D_i$ wins or coexistence	$I_j$ wins	$D_i$ wins	Standoff
(c) $E_i E_j$ mixture	$C_i$	$I_i$	$E_i$	$D_i$
$E_i E_j$	$E_i E_j$ wins	$E_i E_j$ wins	–	$E_i E_j$ wins or $E_i D_i$ standoff

spatial configuration (Supplementary Figs. A.2, A.3, A.9). In the second way (e.g.,  $I_i$  beats  $D_j$ ), an individual of the winning tag-by-strategy type is unable to exploit an individual of the losing type when they are adjacent (both receive a payoff of  $P$ ; Eq. (1)), but, as with the first way, the winning type's average payoff with its other three neighbours is greater than or equal to the losing type's average payoff, again regardless of spatial configuration (Supplementary Figs. A.2, A.3).

Finally,  $E_i$  versus  $E_j$  leads to a unique outcome among pairs of tag-by-strategy types. Both  $E_i$  and  $E_j$  perform worse with others of the same tag than with others of a different tag (respective payoffs  $P$  and  $R$ ; Eq. (1)). Thus, extra-tag cooperators of different tags are mutually invisable at the neighbourhood scale, and readily form mixed-tag aggregations (Supplementary Figs. A.2, A.4, A.13). These aggregations, which can be considered a type of 'meta-strategy', are stable due to negative frequency-dependence: both  $E_i$  and  $E_j$  can spread from rarity in a patch of the other.

In such situations, in order to understand the outcome of spatial interactions it is important to consider the collective properties of such aggregations (van Baalen and Rand, 1998). In this case, the heterogeneous nature of mixed-tag aggregations of  $E$  strategists has the potential to make them strong spatial competitors against the other tag-by-strategy types. Within a patch of same-tagged individuals,  $E$  strategists are able to exploit their more cooperative tag-mates in the same fashion as  $D$  strategists. However, monotypic clusters of  $E$  strategists fare very poorly in patches of different-tagged individuals and are therefore expected to be exploited at patch boundaries. This situation may be mitigated in mixed-tag aggregations, because no matter the tag of an abutting patch, the mixed-tag aggregation always has a local supply of same-tagged  $E$  strategists that can perform well against this abutting patch by defecting. Moreover, within mixed-tag patches of  $E$  strategists, and in contrast to within patches of  $D$  strategists, most interactions are mutually cooperative. Thus, mixed-tag aggregations of  $E$  strategists reap the rewards of cooperation internally, yet may be difficult to exploit along their frontiers, leading to the preliminary prediction that they should be able to hold fast or even overtake multiple other strategies. This provides an interesting link to the theory of pre-biotic evolution in which whole hypercycles resist 'parasite molecules' even when such parasites can outcompete component species of the hypercycle (Boerlijst and Hogeweg, 1991).

The preliminary prediction from above can be refined by considering how two-tag mixed-aggregations of  $E$  strategists perform against the other (monotypic) strategies. Unlike the interactions between pairs of same-tagged strategies (Table 1a) or different-tagged strategies (Table 1b), the outcomes of interactions involving mixed-tag aggregations of  $E$  strategists (e.g.,  $E_iE_j$  for a two-tag system; Table 1c; Appendix B) are difficult to predict analytically. However, simulations show that, in the absence of mutation,  $E_iE_j$  beats  $C_i$  and  $I_i$  regardless of  $r$  (Supplementary Figs. B.1, B.2, B.3). The situation when  $E_iE_j$  meets  $D_i$  is more complicated (Supplementary Figs. B.1, B.4–B.6):  $E_iE_j$  can eliminate  $D_i$  when  $r$  is relatively low, but for increasing  $r$ ,  $D_i$  is able to first persist and then even eliminate  $E_j$  from the mixed-tag aggregation leaving a  $E_i-D_i$  standoff (which, if other tag-by-strategy types are also present, makes  $E_i$  very vulnerable to exploitation; see Table 1b). Therefore, mixed-tag aggregations are predicted to be particularly effective when the cost–benefit ratio of mutual cooperation is low.

### 3.2. Two-tag models with all four strategies

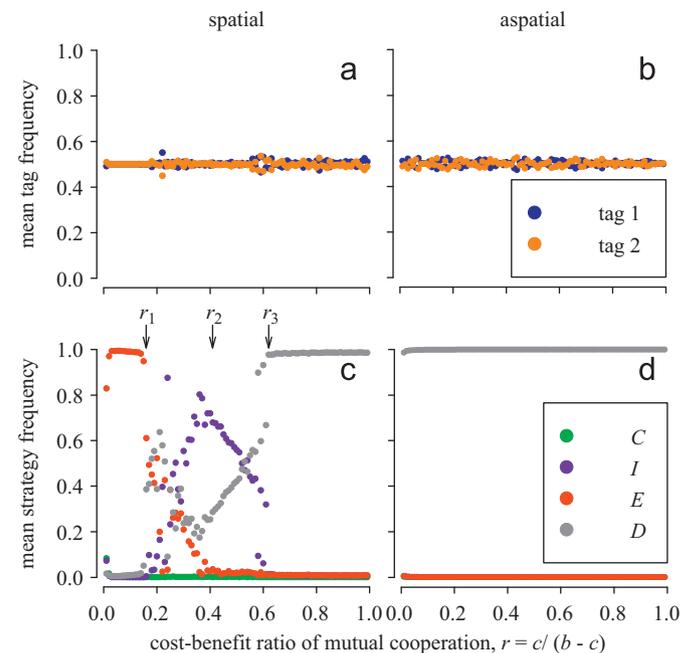
This section gives the full model results for two-tag systems (Tag  $i=1$  and Tag  $j=2$ ). Here, all four strategies are initially present, and mutation allows the potential reappearance of any

tag-by-strategy types that go extinct. To elucidate the role of spatial interactions, the results from the spatial model are contrasted with those from the aspatial model. These results are then interpreted in light of the pairwise strategy dynamics (Section 3.1), which predict what will happen along the borders separating different tag-by-strategy patches (also see Appendix C).

In the aspatial model, which corresponds to a well-mixed population,  $D$  rapidly excludes the other strategies from the population regardless of the cost–benefit ratio,  $r$  (Fig. 1d). In such a situation, tags serve no function and the two tags coexist neutrally (Fig. 1b). Thus, while tag-based cooperation can evolve in unstructured populations when the tag-mutation rate is relatively large compared to the strategy-mutation rate (Antal et al., 2009), this requisite condition does not occur here and so unconditional defection dominates.

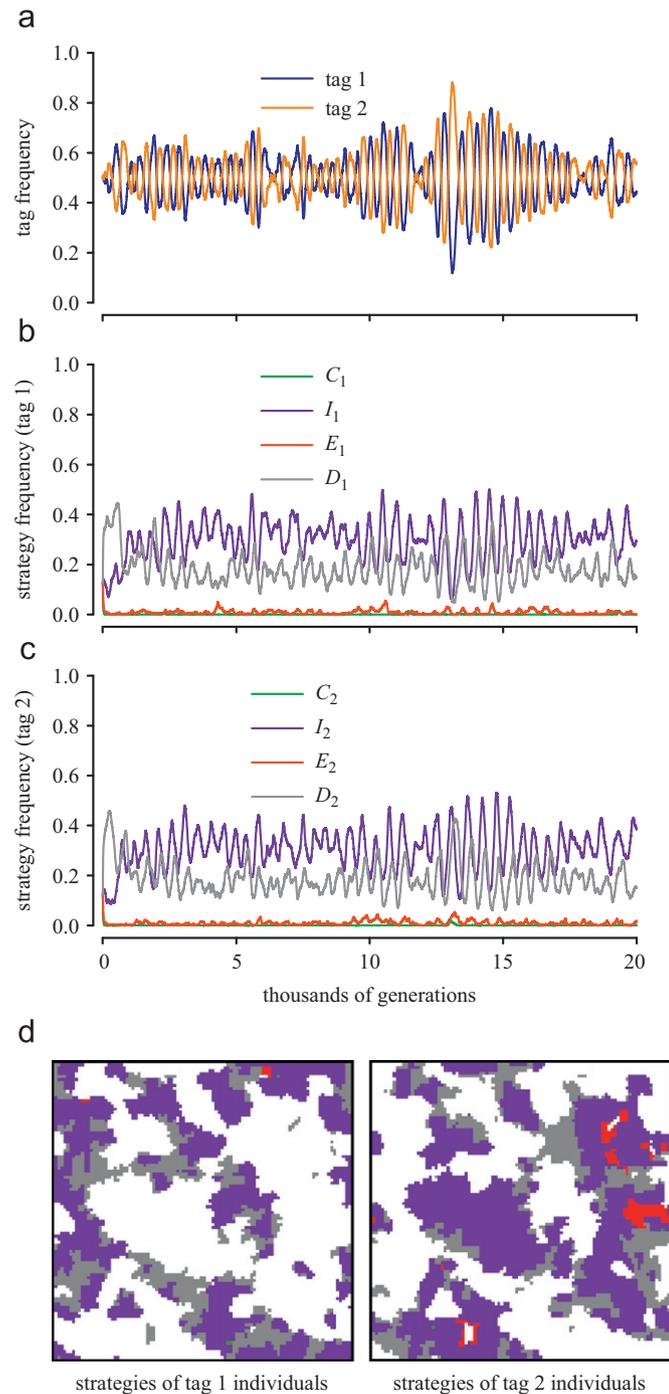
Conversely, in the corresponding spatial model, where interactions take place in local neighbourhoods, conditionally cooperative strategies can evolve. Moreover, there are three critical values of the cost–benefit ratio of mutual cooperation,  $r_1 \approx 0.16$ ,  $r_2 \approx 0.41$ , and  $r_3 \approx 0.62$  (Fig. 1c). These critical values delineate four regions of parameter space, each with qualitatively distinct tag and strategy dynamics (movies illustrating the different dynamical patterns are available in Appendix D).

For  $r > r_3$ , spatial structure cannot rescue cooperation; as in the aspatial model,  $D$  rapidly dominates the population (Fig. 1c; Supplementary Fig. C.4; Supplementary Movie D.4). However, for slightly lower cost–benefit ratios ( $r_2 < r < r_3$ ), the traditional green-beard effect takes hold, allowing  $I$  to dynamically coexist



**Fig. 1.** Space affects tag and strategy coexistence in the two-tag Prisoner's dilemma game. Panels show the average tag or strategy frequencies (the latter cumulative across both tags) between 15- and 20-thousand model generations for populations of 10,000 individuals, cost–benefit ratios  $0 < r < 1$  in increments of 0.01, and a mutation rate  $\mu = 10^{-4}$ . As predicted by evolutionary dynamics (Nowak, 2006a),  $D$  (grey) dominates in the aspatial tag-based Prisoner's dilemma (d), leading to neutral coexistence of Tag 1 (blue) and Tag 2 (orange) (b). In the spatial version, other combinations of strategies can evolve, provided that  $r$  is less than a critical value  $r_3 \approx 0.62$  (c). When  $r$  is greater than  $r_2 \approx 0.41$  but less than  $r_3$ ,  $I$  (purple) and  $D$  coexist. When  $r$  is between  $r_1 \approx 0.16$  and  $r_2$ , then  $I$ ,  $D$ , and  $E$  (red) coexist in evolutionary time. Finally, when  $r < r_1$ ,  $E$  dominates. In no cases does  $C$  (green) evolve. In all cases, Tags 1 and 2 coexist – either neutrally ( $r > r_3$ ) or dynamically ( $r < r_3$ ) (a). Initially there was a random spatial arrangement of tags and strategies; the same long-term patterns emerge when the population starts with only unconditional defectors of a single tag (see Appendix F, Fig. F.1).

with  $D$  (Figs. 1c and 2). This coexistence occurs due to variation in within- and between-tag invasibility which arises when spatial structure promotes the spontaneous formation of local monocultures (Fig. 2d). Specifically, intra-tag cooperators of one



**Fig. 2.** Cyclic coexistence of  $I$  and  $D$  for  $r_2 < r < r_3$ . For relatively high cost–benefit ratios between  $r_2$  and  $r_3$  (parameter values in this example:  $r=0.45$ ,  $\mu=10^{-4}$ ), and starting from a random spatial arrangement of tags and strategies, monocultures of  $I_1$ ,  $D_1$ ,  $I_2$ , and  $D_2$  form spontaneously and invade one another in a non-transitive fashion, due to the fact that  $D_1$  and  $D_2$  can invade  $I_1$  and  $I_2$ , respectively, but  $I_1$  and  $I_2$  can invade  $D_2$  and  $D_1$ , respectively (Supplementary Fig. C.3) (Axelrod et al., 2004; Hammond and Axelrod, 2006a; Traulsen, 2008). This produces cyclic dynamics of tags (a) and strategies (b, c). (d) shows the spatial arrangement of strategies after 20,000 generations for the same model run for Tag 1 (left) and Tag 2 (right). Colours are as in Fig. 1 and (b and c); white areas are occupied by the other column's tag, and hence are complementary. Supplementary Movie D.3 shows the evolution of the lattice for the same parameter values.

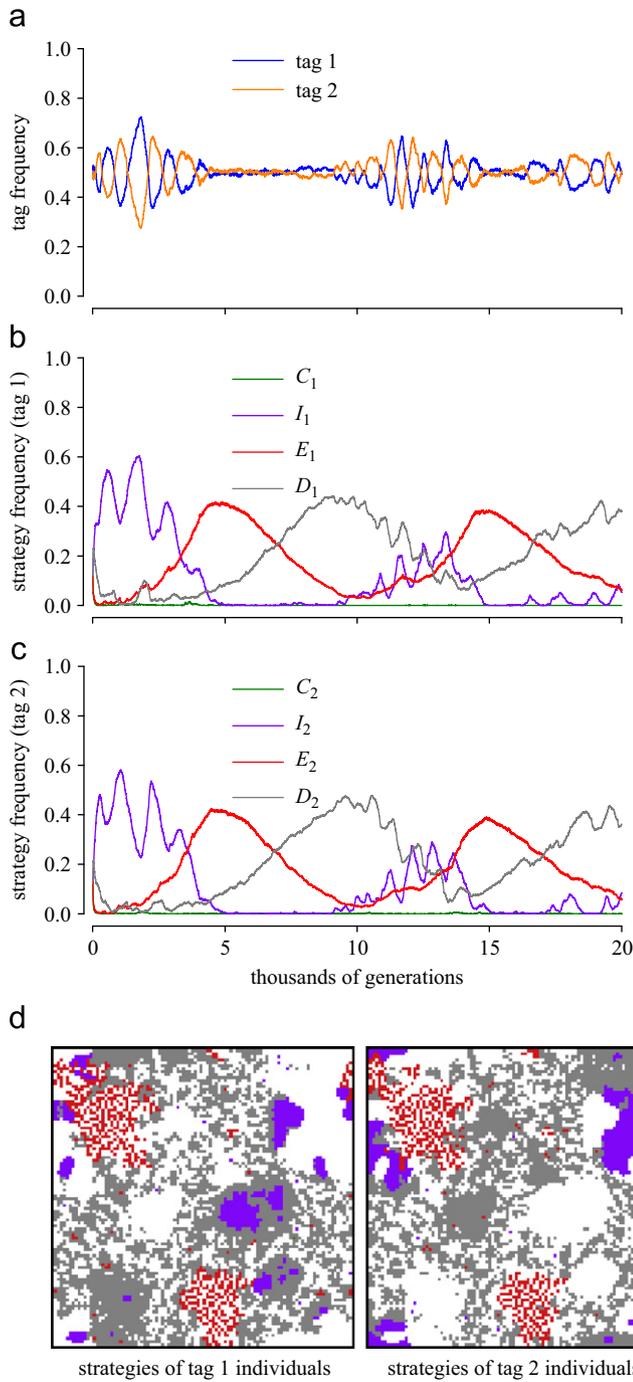
tag can invade unconditional defectors of the other tag, whilst they are simultaneously invaded by unconditional defectors of their own tag (Hammond and Axelrod, 2006a), resulting in a non-transitive invasion loop whereby  $I_1 \rightarrow D_1 \rightarrow I_2 \rightarrow D_2 \rightarrow I_1$  (Supplementary Fig. C.3). This causes cyclic dynamics in which  $I_1$  and  $I_2$  are a half-cycle out of phase,  $D_1$  and  $D_2$  are a half-cycle out of phase, and each  $D$  lags a partial cycle behind the corresponding  $I$  of the same tag (Fig. 2b and c; Supplementary Movie D.3). This result reinforces the finding that the green-beard effect can lead to cooperation even in the presence of unconditional defectors (Axelrod et al., 2004; Hammond and Axelrod, 2006a, 2006b; Jansen and van Baalen, 2006; Masuda and Ohtsuki, 2007; Traulsen and Nowak, 2007; Traulsen, 2008; Antal et al., 2009; Shultz et al., 2009). Additionally, the cyclic replacement of strategies is similar to the cyclic replacement of honest and dishonest communicators found by van Baalen and Jansen in their analysis of the evolution of signals and their meanings (van Baalen and Jansen, 2003).

For  $r_1 < r < r_2$ , the dynamics become more complicated as  $E$  coexists in evolutionary time with  $I$  and  $D$  (Fig. 1c). Due to mutation and spatial stochasticity, mixed-tag aggregations of  $E_1$  and  $E_2$  occasionally form and grow. These aggregations are characterised by fine-grain spatial partitioning of the two tags (Fig. 3d; Supplementary Figs. A.4, A.13), so that individuals within such an aggregation enjoy a relatively high degree of cooperation. For this region of parameter space, these coalitions of extra-tag cooperators can invade monocultures of  $I_1$  and  $I_2$ , yet are themselves invaded by monocultures of  $D_1$  and  $D_2$  (Supplementary Fig. C.2). This results in non-transitive invasibility and long-term cyclic replacement of  $I$ ,  $E$ , and  $D$ , all of which is superimposed over the shorter-term cycle of  $I_1$ ,  $D_1$ ,  $I_2$ , and  $D_2$  (Fig. 3b and c; Supplementary Movie D.2). Thus, the same green-beard effect that promotes within-group cooperation can also lead to periodic outbreaks of traitorous between-group cooperation.

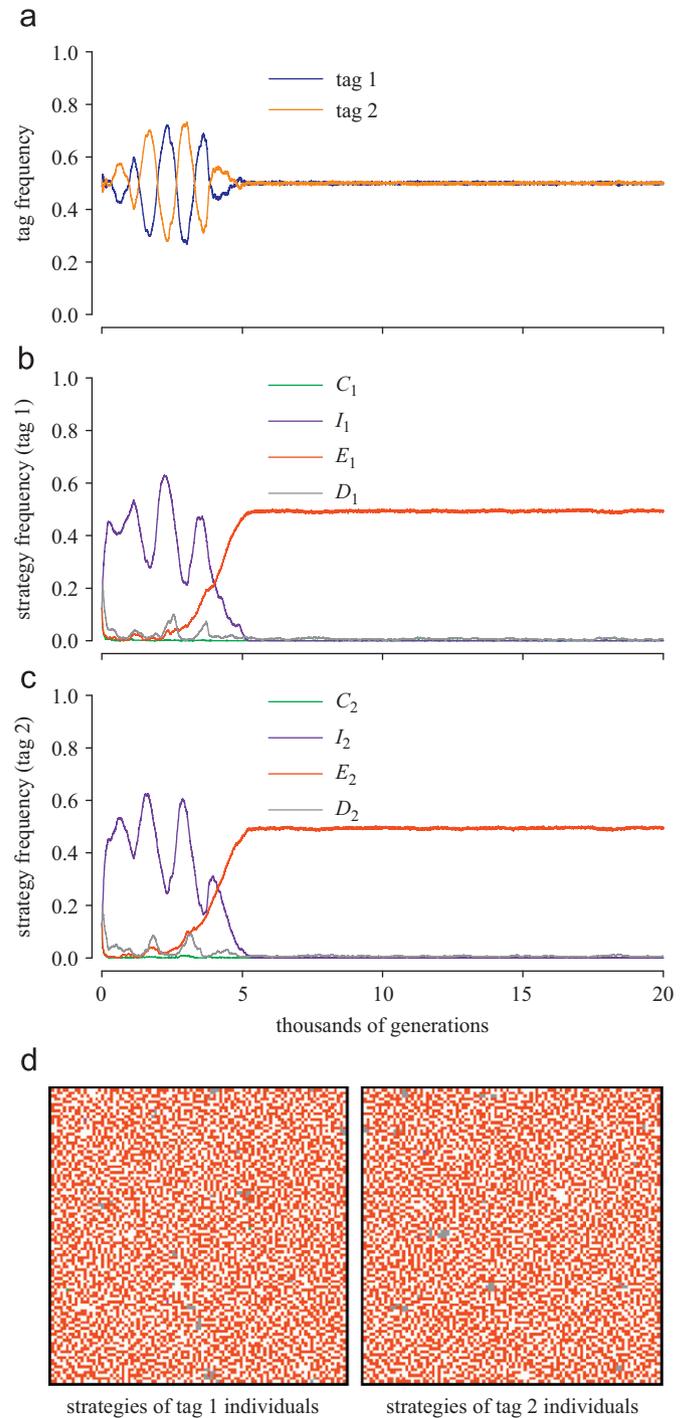
Finally, at low cost–benefit ratios ( $r < r_1$ ), mixed-tag aggregations of extra-tag cooperators dominate the population (Fig. 1c). The situation is similar to when  $r_1 < r < r_2$ , except that for lower values of  $r$ , these finely partitioned, mixed-tag groups of extra-tag cooperators are invulnerable to invasion by unconditional defectors, and therefore represent the end point of the invasion dynamics (Fig. 4d; Supplementary Fig. C.1). Therefore, after transient cyclic dominance by  $I_1$ ,  $D_1$ ,  $I_2$ , and  $D_2$ , a group composed of  $E_1$  and  $E_2$  takes over the population (Fig. 4b and c; Supplementary Movie D.1; Supplementary Fig. D1). In this region of parameter space, traitorousness in the form of extra-tag cooperation is the quasi-stable strategy.

In the absence of other mechanisms, population structure is imperative to allow conditionally cooperative strategies (including traitorousness) to evolve. Here, the population structure is spatial in nature. Neighbourhood interactions allow local monocultures of  $I$  strategists and local mixed-tag aggregations of  $E$  strategists to form, causing conditional cooperators to interact with one another vastly more frequently than predicted by their relative abundances alone. Further, because dispersal is local in the spatial model, interacting individuals typically share a more recent common ancestor compared to interacting individuals in the aspatial model. As such, the 'spatial network selection' mechanism presented here can be cast in terms of kin selection, in which close kin have the potential to interact preferentially due to their spatial proximity (Taylor et al., 2007).

In addition to the importance of population structure, the existence of mutation is critical for the evolution of the extra-tag cooperator strategy in particular (Appendix E). Mutation allows mixed-tag groups of  $E$  strategists to form and take hold in a population that is otherwise predominantly controlled by other



**Fig. 3.** Non-transitive spatial invasion of  $I$ ,  $E$ , and  $D$  for  $r_1 < r < r_2$ . For intermediate cost–benefit ratios (parameter values in this example:  $r=0.19$ ,  $\mu=10^{-4}$ ), and starting from a random spatial arrangement of tags and strategies,  $I$ ,  $D$ , and  $E$  spontaneously form aggregations that can invade one another in a non-transitive fashion allowing for their coexistence over evolutionary time (Supplementary Fig. C.2). (b, c) First,  $I_1$ ,  $I_2$ ,  $D_1$ , and  $D_2$  fluctuate cyclically as in Fig. 2. Eventually, mutation and spatial stochasticity allow a mixed-tag aggregation of  $E_1$  and  $E_2$  to form which can invade the  $I$ -dominated lattice. However, this mixed aggregation is itself vulnerable to invasion by  $D_1$  and  $D_2$ . The resulting  $D$ -dominated lattice can be invaded by  $I_1$  and  $I_2$ , closing the non-transitive invasion loop. Meanwhile, the tag frequencies cycle when  $I$  and  $D$  cycle, but are relatively constant otherwise (a). (d) shows the spatial arrangement of strategies after 20,000 generations for the same model run for Tag 1 (left) and Tag 2 (right). Colours are as in Fig. 1 and (b and c). Supplementary Movie D.2 shows the evolution of the lattice for the same parameter values.



**Fig. 4.** Dominance of traitorousness (extra-tag cooperation) for  $r < r_1$ . For low cost–benefit ratios (parameter values in this example:  $r=0.12$ ,  $\mu=10^{-4}$ ), and starting from a random spatial arrangement of tags and strategies, mixed-tag groups of  $E_1$  and  $E_2$  eventually come to dominate the population. (b, c) Evolution initially occurs as in Fig. 3; however, for this low value of  $r$ , mixed-tag aggregations of  $E_1$  and  $E_2$  are invulnerable to invasion by the other strategies and thus come to dominate (Supplementary Fig. C.1). The tag frequencies cycle during the period of transient cyclic dominance by  $I$  and  $D$ , but become constant once the mixed-tag aggregation of  $E$  takes over (a). (d) shows the spatial arrangement of strategies after 20,000 generations for the same model run for Tag 1 (left) and Tag 2 (right). Colours are as in Fig. 1 and (b and c). Supplementary Movie D.1 shows the evolution of the lattice for the same parameter values.

strategies. When mutation is omitted from the model, extra-tag cooperation only rarely comes to dominate the population (Supplementary Fig. E.1). Likewise, when the mutation rate is very high (e.g., 1%), mixed-tag groups of *E* strategists are undermined by *D* strategists before they can grow (Supplementary Fig. E.2). Thus, the model predicts that the evolution of traitorousness requires mutation to be present, but relatively rare, as is typical in most biological systems.

### 3.3. Three-, four-, and five-tag models with all four strategies

Intriguingly, for similar spatial models with more than two tags, extra-tag cooperation does not evolve. Rather, *I* strategists dominate at relatively low values of *r* with *D* strategists becoming increasingly frequent as *r* increases (results for three–five-tag models are provided in Appendix G, Supplementary Figs. G.1–G.3).

There are two main reasons why the two-tag model and the multi-tag models differ so dramatically. First, there is the matter of mutant limitation. In the two-tag model, it often takes many generations before mutation and spatial stochasticity allow an *E*<sub>1</sub> and an *E*<sub>2</sub> patch to overlap sufficiently to instigate a mixed-tag aggregation (e.g., Supplementary Movie D.1). In the three-tag model, the problem is greatly compounded, because patches of *E*<sub>1</sub>, *E*<sub>2</sub>, and *E*<sub>3</sub> must overlap before a mixed-tag aggregation can have any chance of sustained growth (note that a two-tag aggregation is ineffectual in a three-tag system: a mixed-tag aggregation of *E*<sub>1</sub> and *E*<sub>2</sub> would be easily exploited by a patch of *I*<sub>3</sub> or *D*<sub>3</sub>). However, mutant limitation is not the only reason for the lack of success of *E* strategists in multi-tag models; indeed, in the presence of rare mutation, three-tag models initiated with only *E*<sub>1</sub>, *E*<sub>2</sub>, and *E*<sub>3</sub> are quickly invaded and taken over by other strategies, even for low values of *r* (Supplementary Fig. G.4). The second reason why mixed-tag aggregations of *E* strategists are unsustainable in multi-tag systems is the dilution of each tag within the aggregation. This can best be understood in contrast to a two-tag system. In a two-tag system at sufficiently low *r*, a mixed-tag aggregation of *E*<sub>1</sub> and *E*<sub>2</sub> cannot be overtaken by *I*<sub>1</sub> (for example) because even though *I*<sub>1</sub> can invade *E*<sub>2</sub>, there is always a local supply of *E*<sub>1</sub> available to resist *I*<sub>1</sub>'s further incursion (e.g., Supplementary Fig. D.1). Although three-tag aggregations of *E* strategists perform approximately as well against other (monotypic) tag-by-strategy types as their two-tag counterparts (compare Supplementary Figs. B.1 and G.5; also see Supplementary Figs. G.6–G.10), such three-tag aggregations are too dilute to resist all three tags of *I* strategists simultaneously. Naturally, the situation is exacerbated in four-, five-, or more-tag systems.

Thus, the traditional green-beard effect operates efficiently in spatially structured populations, even in the presence of traitorous extra-tag cooperators, provided the cost–benefit ratio of mutual cooperation is not too high and provided there is a diversity of tags – i.e., beard colours – in the population.

### 3.4. More sophisticated strategies in chromodiverse populations

One avenue for future research is the examination of more sophisticated strategies. In particular, in populations with more than two tags, there are additional fixed strategies beyond *C*, *I*, *E*, and *D*. For example, in a three-tag system, one strategy adopted by Tag 1 individuals might be to cooperate only with those bearing Tag 2, while defecting against Tags 1 and 3. When paired with a 'complementary' tag-by-strategy type (e.g., Tag 2 individuals that cooperate only with those bearing Tag 1 and no others), it is possible that mixed-tag aggregations will arise that can avoid the problems associated with mixed-tag *E* strategists discussed in Section 3.3 (similar to two-tag hypercycles that resist parasites (Boerlijst and Hogeweg, 1991)). However, even if such mixed-tag

meta-strategies ultimately prove successful, the *C–I–E–D* scheme explored here is still worth pursuing for at least two reasons. First, it provides a natural comparison with the two-tag case in which more sophisticated fixed strategies do not occur. Second, more sophisticated strategies may require correspondingly more sophisticated sensory systems to distinguish between multiple different tags (whereas the conditional strategies *I* and *E* can operate through the simpler mechanism of self-reference). Thus, depending on the biological system in question, more sophisticated conditional strategies might not be available in practice.

### 3.5. Conclusion

Although it is usually not thought of as such, traitorousness is a type of cooperation. Here I show that in structured populations with two recognisable phenotypic tags, evolutionary game dynamics predicts that traitors can flourish and even come to dominate populations. By tuning the cost–benefit ratio of mutual cooperation, the same basic premises of the green-beard effect that were initially proposed to explain intra-tag cooperation also inevitably lead to traitorous extra-tag cooperation in two-tag systems. Interestingly, however, the green-beard mechanism of intra-tag cooperation appears to be relatively immune to traitorousness in systems with a greater number of tags/beard colours, making tag-based cooperation more robust than previously appreciated (with the caveat that non-*C*, *-I*, *-E*, and *-D* strategy-variants will require further analysis in chromodiverse populations). More broadly, these results emphasise that in structured populations, strategy evolution should be considered at the level of the cluster, not just the individual (van Baalen and Rand, 1998).

### Acknowledgements

I. Cuthill, T. Day, K. Kwok, K. Judge, B. Schamp, T. Sherratt, M. van Baalen, and an anonymous reviewer provided helpful comments on the article. I thank the Natural Sciences and Engineering Research Council (Canada) and the University of Lethbridge for funding.

### Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2011.07.023.

### References

- Antal, T., Ohtsuki, H., Wakeley, J., Taylor, P.D., Nowak, M.A., 2009. Evolution of cooperation by phenotypic similarity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8597–8600.
- Axelrod, R., 2006. *The Evolution of Cooperation*, revised ed.. Basic Books, Cambridge, USA.
- Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. *Science* 211, 1390–1396.
- Axelrod, R., Hammond, R.A., Grafen, A., 2004. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution* 58, 1833–1838.
- Boerlijst, M.C., Hogeweg, P., 1991. Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D* 48, 17–28.
- Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press, Oxford.
- Dawkins, R., 1982. *The Extended Phenotype*. Oxford University Press, Oxford.
- Doebeli, M., Hauert, C., 2005. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecol. Lett.* 8, 748–766.
- Durrett, R., Levin, S.A., 1994. Stochastic spatial models: a user's guide to ecological applications. *Philos. Trans. R. Soc. Lond. B* 343, 329–350.
- Gardner, A., West, S.A., 2009. Greenbeards. *Evolution* 64, 25–38.
- Hamilton, W.D., 1964. The genetical evolution of social behaviour: I and II. *J. Theor. Biol.* 7, 1–52.

- Hammond, R.A., Axelrod, R., 2006a. The evolution of ethnocentrism. *J. Conflict Resol.* 50, 926–936.
- Hammond, R.A., Axelrod, R., 2006b. Evolution of contingent altruism when cooperation is expensive. *Theor. Popul. Biol.* 69, 333–338.
- Hauert, C., Doebeli, M., 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428, 643–646.
- Jansen, V.A.A., van Baalen, M., 2006. Altruism through beard chromodynamics. *Nature* 440, 663–666.
- Keller, L., Ross, K.G., 1998. Selfish genes: a green beard in the red fire ant. *Nature* 394, 573–575.
- Lehmann, L., Feldman, M.W., Rousset, F., 2009. On the evolution of harming and recognition in finite panmictic and infinite structured populations. *Evolution* 63, 2896–2913.
- Lieberman, E., Hauert, C., Nowak, M.A., 2005. Evolutionary dynamics on graphs. *Nature* 433, 312–316.
- Masuda, N., Ohtsuki, H., 2007. Tag-based indirect reciprocity by incomplete social information. *Proc. R. Soc. B* 274, 689–695.
- Maynard Smith, J., 1982. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, UK.
- Maynard Smith, J., Szathmáry, E., 1997. *The Major Transitions of Evolution*. Oxford University Press, New York.
- Nowak, M.A., 2006a. Five rules for the evolution of cooperation. *Science* 314, 1560–1563.
- Nowak, M.A., 2006b. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, Cambridge.
- Nowak, M.A., May, R.M., 1992. Evolutionary games and spatial chaos. *Nature* 359, 826–829.
- Nowak, M.A., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577.
- Nowak, M.A., Sigmund, K., 2004. Evolutionary dynamics of biological games. *Science* 303, 793–799.
- Queller, D.D., Ponte, E., Bozzaro, S., Straussmann, J.E., 2003. Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. *Science* 299, 105–106.
- Rand, D.G., Armao IV, J.J., Nakamaru, M., Ohtsuki, H., 2010. Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* 265, 624–632.
- Riolo, R.L., Cohen, M.D., Axelrod, R., 2001. Evolution of cooperation without reciprocity. *Nature* 414, 441–443.
- Roberts, G., Sherratt, T.N., 2002. Does similarity breed cooperation? *Nature* 418, 499–500.
- Sherratt, T.N., Wilkinson, D.M., 2009. *Big Questions in Ecology and Evolution*. Oxford University Press, Oxford.
- Shultz, T.R., Hartshorn, M., Kaznatcheev, A., 2009. Why is ethnocentrism more common than humanitarianism?. In: Taatgen, N.A., van Rijn, H. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Amsterdam, pp. 2100–2105.
- Smukalla, S., Caldara, M., Pochet, N., Beauvais, A., Gaudagnini, S., Yan, C., Vences, M.D., Jansen, A., Prevost, M.C., Latgé, J.-P., Fink, G.R., Foster, K.R., Verstrepen, K.J., 2008. *FLO1* is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* 135, 726–737.
- Taylor, P.D., Day, T., Wild, G., 2007. Evolution of cooperation in a finite homogeneous graph. *Nature* 447, 469–472.
- Traulsen, A., 2008. Mechanisms for similarity based cooperation. *Eur. Phys. J. B* 63, 363–371.
- Traulsen, A., Nowak, M.A., 2006. Evolution of cooperation by multilevel selection. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10952–10955.
- Traulsen, A., Nowak, M.A., 2007. Chromodynamics of cooperation in finite populations. *PLoS ONE* 3, e270.
- Traulsen, A., Schuster, H.G., 2003. Minimal model for tag-based cooperation. *Phys. Rev. E* 68, 046129.
- van Baalen, M., Jansen, V.A.A., 2003. Common language or Tower of Babel? On the evolutionary dynamics of signals and their meanings. *Proc. R. Soc. B* 270, 69–76.
- van Baalen, M., Rand, D.A., 1998. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* 193, 631–648.
- West, S.A., Gardner, A., 2010. Altruism, spite, and greenbeards. *Science* 327, 1341–1344.

# Green-beard effect predicts the evolution of traitorousness in the two-tag Prisoner's Dilemma

Electronic supplementary material

Robert A. Laird

*Department of Biological Sciences, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada*

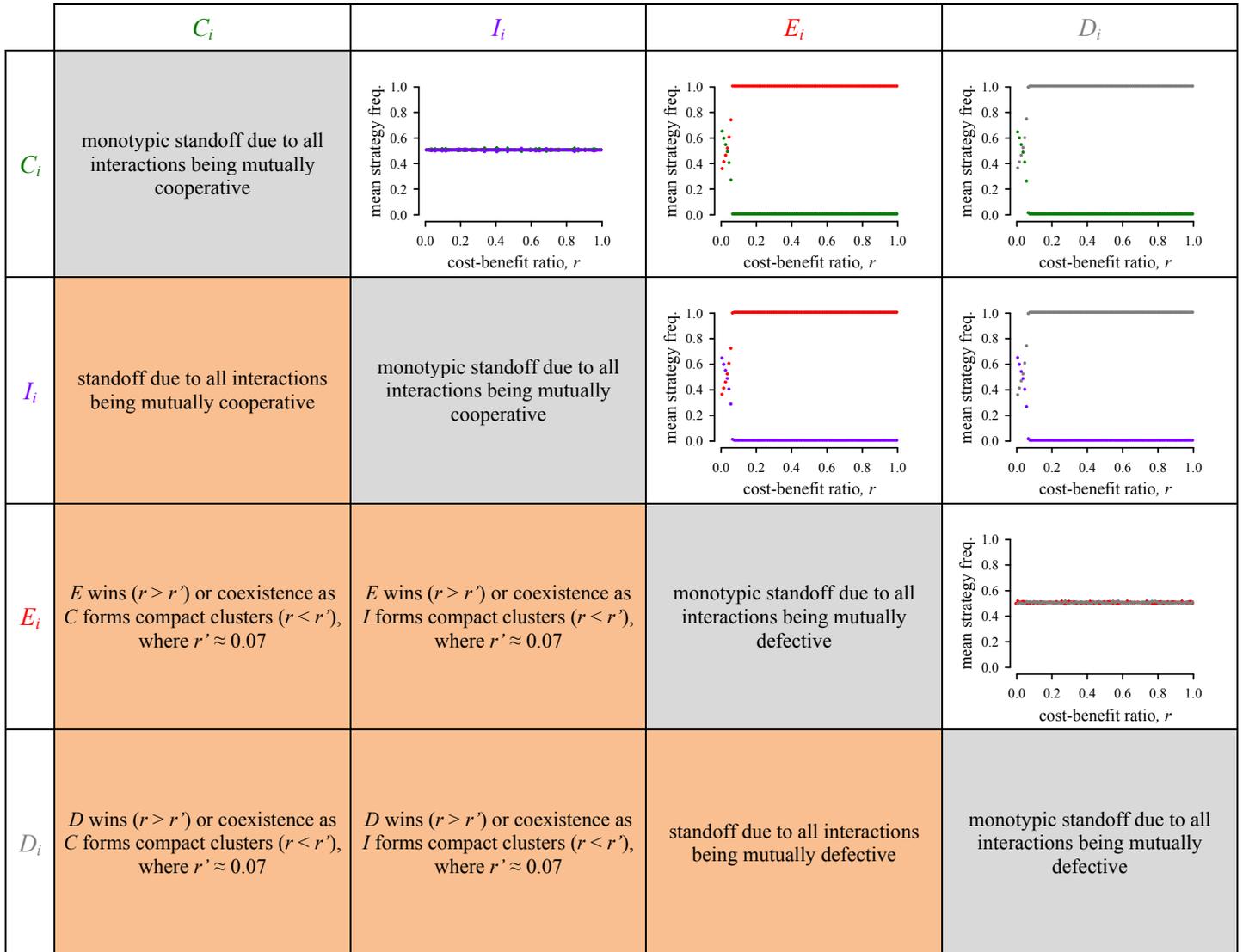
## Contents

<b>Appendix A: <i>Pairwise strategy outcomes in spatially structured populations in the absence of mutation</i></b>	<b>3</b>
Fig. A.1. Spatial invasion outcomes when only two strategies of the same tag are present	3
Fig. A.2. Spatial invasion outcomes when only two strategies of different tags are present	4
Fig. A.3. Interactions between pairs of tag-by-strategy types in which one member of the pair wins outright in the spatial arena	5
Fig. A.4. Fine-grain spatial partitioning by mixed-tag aggregations of extra-tag cooperators	5
Fig. A.5. Snapshots of spatial lattice model for $C_i$ versus $C_j$ ( $\mu = 0, r = 0.50$ )	7
Fig. A.6. Snapshots of spatial lattice model for $C_i$ versus $D_j$ ( $\mu = 0, r = 0.06$ )	8
Fig. A.7. Snapshots of spatial lattice model for $C_i$ versus $D_j$ ( $\mu = 0, r = 0.08$ )	9
Fig. A.8. Snapshots of spatial lattice model for $D_i$ versus $D_j$ ( $\mu = 0, r = 0.50$ )	10
Fig. A.9. Snapshots of spatial lattice model for $C_i$ versus $I_j$ ( $\mu = 0, r = 0.50$ )	11
Fig. A.10. Snapshots of spatial lattice model for $C_i$ versus $E_j$ ( $\mu = 0, r = 0.50$ )	12
Fig. A.11. Snapshots of spatial lattice model for $I_i$ versus $I_j$ ( $\mu = 0, r = 0.32$ )	13
Fig. A.12. Snapshots of spatial lattice model for $I_i$ versus $I_j$ ( $\mu = 0, r = 0.31$ )	14
Fig. A.13. Snapshots of spatial lattice model for $E_i$ versus $E_i$ ( $\mu = 0, r = 0.50$ )	15
<b>Appendix B: <i>Outcomes involving <math>E_iE_j</math> aggregations in spatially structured populations with no mutation</i></b>	<b>16</b>
Fig. B.1. $E_iE_j$ versus $C_i, I_i$ , or $D_i$	16
Fig. B.2. Snapshots of spatial lattice model for $E_iE_j$ versus $C_i$ ( $\mu = 0, r = 0.50$ )	17
Fig. B.3. Snapshots of spatial lattice model for $E_iE_j$ versus $I_i$ ( $\mu = 0, r = 0.50$ )	18
Fig. B.4. Snapshots of spatial lattice model for $E_iE_j$ versus $D_i$ ( $\mu = 0, r = 0.26$ )	19
Fig. B.5. Snapshots of spatial lattice model for $E_iE_j$ versus $D_i$ ( $\mu = 0, r = 0.29$ )	20
Fig. B.6. Snapshots of spatial lattice model for $E_iE_j$ versus $D_i$ ( $\mu = 0, r = 0.38$ )	21
<b>Appendix C: <i>Graphical interpretation of outcomes of the spatial, two-tag Prisoner's Dilemma game</i></b>	<b>22</b>
Fig. C.1. Two-tag Prisoner's Dilemma for $r < r_1$	22
Fig. C.2. Two-tag Prisoner's Dilemma for $r_1 < r < r_2$	23
Fig. C.3. Two-tag Prisoner's Dilemma for $r_2 < r < r_3$	23
Fig. C.4. Two-tag Prisoner's Dilemma for $r > r_3$	24

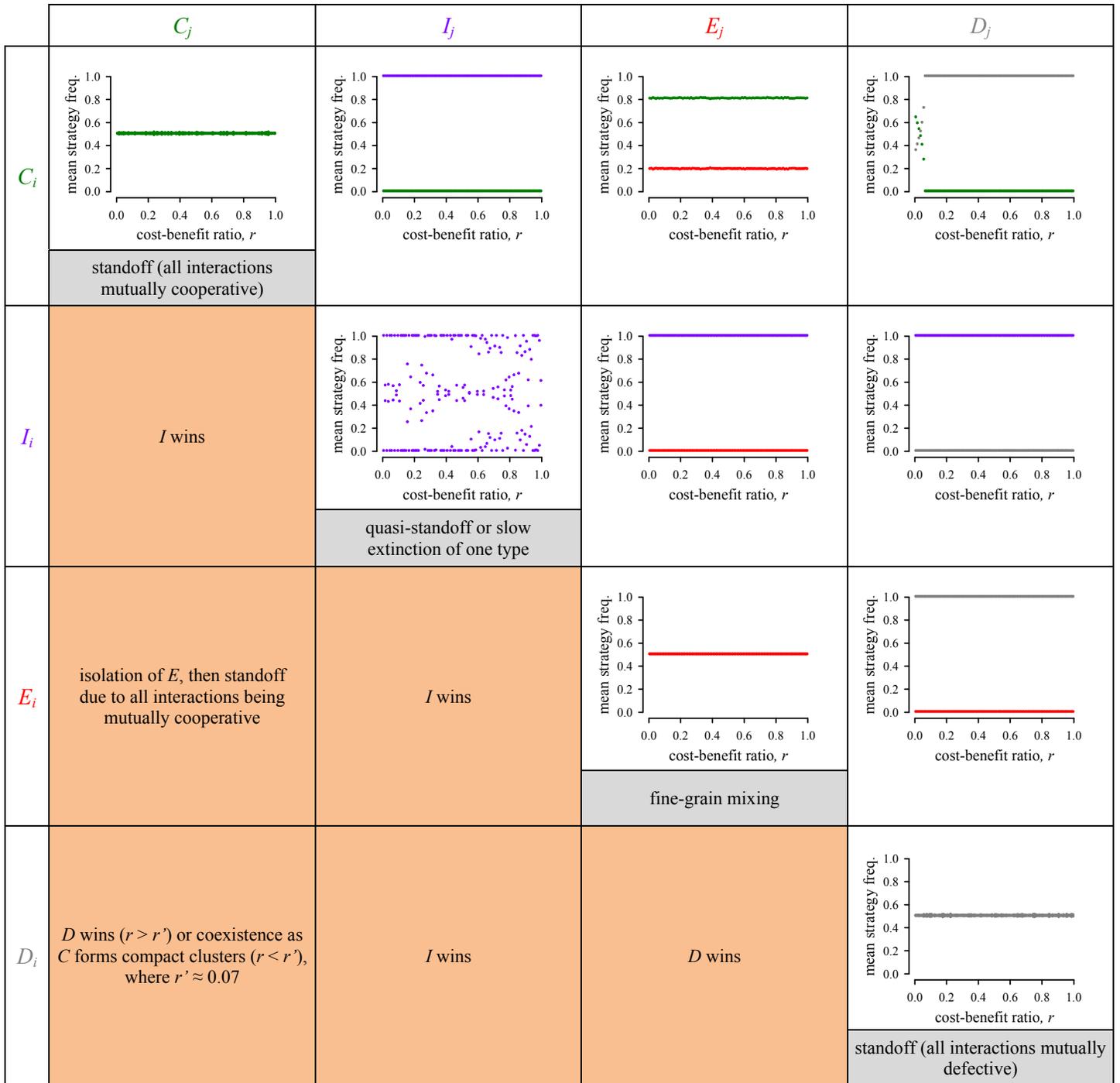
Appendix D: <i>Example model runs for the spatial, two-tag Prisoner's Dilemma game</i>	25
Table D.1. Movies of example model runs for the spatial, two-tag Prisoner's Dilemma game	25
Fig. D.1. Typical invasion sequence of a mixed-tag aggregation of extra-tag cooperators	26
Appendix E: <i>Mutation and the evolution of traitorousness in two-tag populations</i>	27
Fig. E.1. Mutation and the evolution of traitorousness: Low versus no mutation	27
Fig. E.2. Mutation and the evolution of traitorousness: Low versus high mutation	28
Appendix F: <i>Different starting conditions in two-tag populations</i>	29
Fig. F.1. Different starting conditions: 'Mixed' versus 'Ancestral'	29
Appendix G: <i>Three-, four-, and five-tag models</i>	30
Fig. G.1. Three-tag models	30
Fig. G.2. Four-tag models	31
Fig. G.3. Five-tag models	31
Fig. G.4. Three-tag models, starting with only $E$ -strategists	32
Fig. G.5. $E_iE_jE_k$ versus $C_i$ , $I_i$ , or $D_i$	32
Fig. G.6. Snapshots of spatial lattice model for $E_iE_jE_k$ versus $C_i$ ( $\mu = 0$ , $r = 0.50$ )	34
Fig. G.7. Snapshots of spatial lattice model for $E_iE_jE_k$ versus $I_i$ ( $\mu = 0$ , $r = 0.50$ )	35
Fig. G.8. Snapshots of spatial lattice model for $E_iE_jE_k$ versus $D_i$ ( $\mu = 0$ , $r = 0.09$ )	36
Fig. G.9. Snapshots of spatial lattice model for $E_iE_jE_k$ versus $D_i$ ( $\mu = 0$ , $r = 0.15$ )	37
Fig. G.10. Snapshots of spatial lattice model for $E_iE_jE_k$ versus $D_i$ ( $\mu = 0$ , $r = 0.38$ )	38

## Appendix A: Pairwise strategy dynamics in spatially structured populations in the absence of mutation

In lattice models of spatial competition, invasion and replacement occur along borders between pairs of patches. Therefore, it is instructive to consider spatial interactions between pairs of types (in the absence of mutation). These situations are sufficiently simple that the outcome of competition can be deduced *a priori*, either by analysis or by reference to previous work (see §3.1, main text). Here, these results are confirmed by simulations. Figure A.1 gives the spatial invasion outcomes when only two strategies of *the same tag* are present. Figure A.2 gives the spatial invasion outcomes when only two strategies of *different tags* are present. In both cases, the lattices started with random spatial configurations.

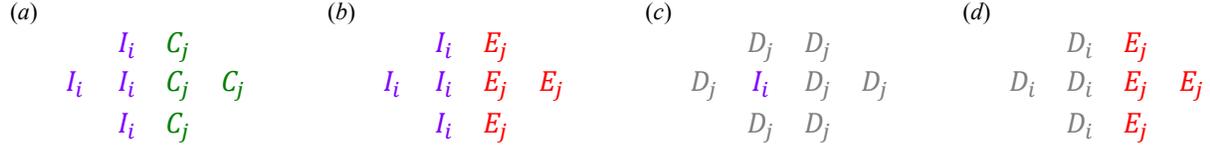


**Fig. A.1.** Spatial invasion outcomes when only two strategies of the *same tag* are present (row versus column) and there is no mutation. This figure confirms through simulations the analytical results presented in Table 1a and §3.1 in the main text. Above the diagonal, each panel shows the average strategy frequency between 15- and 20-thousand generations of the spatial model, for populations of 10,000 individuals, and cost-benefit ratios of  $0 < r < 1$  in increments of 0.01. Initially there was a random arrangement of strategies. Line colours correspond to the colours of the row and column headers. Below the diagonal (shaded in orange), interaction symmetry means that the outcomes are mirror images of those above the diagonal; therefore, rather than repeating the panels, a description of the outcome is shown instead (corresponding to a more detailed version of Table 1a). Along the diagonal (shaded in grey) are the scenarios that only have one type of individual and are therefore result in monotypic standoffs.



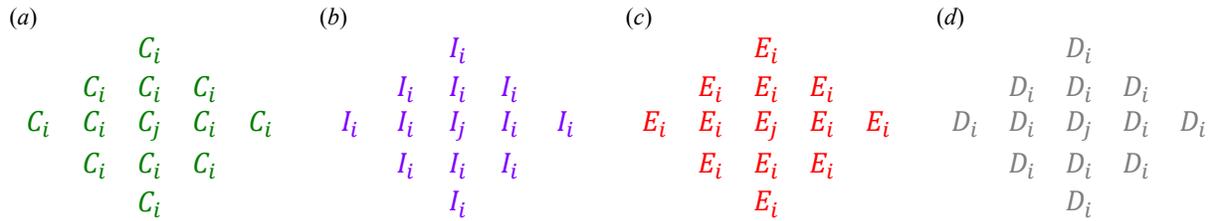
**Fig. A.2.** Spatial invasion outcomes when only two strategies of *different tags* are present (row versus column) and there is no mutation. This figure confirms through simulations the analytical results presented in table 1b and §3.1 in the main text. Above the diagonal, each panel shows the average strategy frequency between 15- and 20-thousand generations of the spatial model, for populations of 10,000 individuals, and cost-benefit ratios of  $0 < r < 1$  in increments of 0.01. Initially there was a random arrangement of strategies. Line colours correspond to the colours of the row and column headers (note that along the diagonal, where the strategies (but not the tags) of the two types are the same, it is not possible to tell apart Tag  $i$  and Tag  $j$ ; however, the distinction is unimportant because the tag names are arbitrary). Below the diagonal (shaded in orange), interaction symmetry means that the outcomes are mirror images of those above the diagonal; therefore, rather than repeating the panels, a description of the outcome is shown instead (corresponding to a more detailed version of Table 1b). A description for the four same-strategy/different-tag scenarios is also provided along the diagonal (shaded in grey).

The four cases where one tag wins outright, no matter the value of the cost-benefit ratio of mutual cooperation (see Fig. A.2), are somewhat more difficult to understand than the other cases, and are explained in Fig. A.3.



**Fig. A.3.** Pairs of tag-by-strategy types in which one member of the pair wins outright in the spatial arena. In each case, the focal interaction is between the middle two individuals, with the individual on the left emerging victorious. Also in each case, the spatial configuration chosen is an instance of one that gives the *lowest* average payoff to the winning tag-by-strategy type and the *greatest* average payoff to the losing tag-by-strategy type, on the grounds that if the winner's payoff exceeds the loser's payoff in the least advantageous spatial configuration, it must also exceed the loser's payoff in all the other possible spatial configurations of the two types. (a)  $I_i$  versus  $C_j$ : The interaction between the focal  $I_i$  and the focal  $C_j$  results in payoffs of  $T$  and  $S$ , respectively. The focal  $I_i$ 's other three payoffs, and the focal  $C_j$ 's other three payoffs, are all  $R$ . Therefore,  $I_i$  has a greater average payoff and overtakes  $C_j$ . (b)  $I_i$  versus  $E_j$ : The interaction between the focal  $I_i$  and the focal  $E_j$  results in payoffs of  $T$  and  $S$ , respectively. The focal  $I_i$ 's other three payoffs are  $R$ ; the focal  $E_j$ 's other three payoffs are  $P$ . Therefore,  $I_i$  has a greater average payoff and overtakes  $E_j$ . (c)  $I_i$  versus  $D_j$ : The interaction between the focal  $I_i$  and the focal  $D_j$  results in a payoff of  $P$  to both participants. The focal  $I_i$ 's other three payoffs, and the focal  $D_j$ 's other three payoffs, are all  $P$ . Thus, in this specific case, there is a standoff; however, if a cluster of  $I_i$  of any size forms,  $I_i$  will have a greater payoff and will overtake  $D_j$ . (d)  $D_i$  versus  $E_j$ : The interaction between the focal  $D_i$  and the focal  $E_j$  results in payoffs of  $T$  and  $S$ , respectively. The focal  $D_i$ 's other three payoffs, and the focal  $E_j$ 's other three payoffs, are all  $P$ . Therefore,  $D_i$  has a greater payoff and overtakes  $E_j$ .

Unlike non-tag-mates of the other three strategies, extra-tag cooperators bearing different tags can exploit one another via negative frequency dependence, and are therefore mutually invisable at the neighbourhood scale. This leads to fine-grain spatial partitioning by  $E$ -strategists of different tags (Fig. A.4; also see Figs. A.13, D.1, the movies defined in Appendix D, and Figs. 3d and 4d in the main text, for examples of this partitioning).



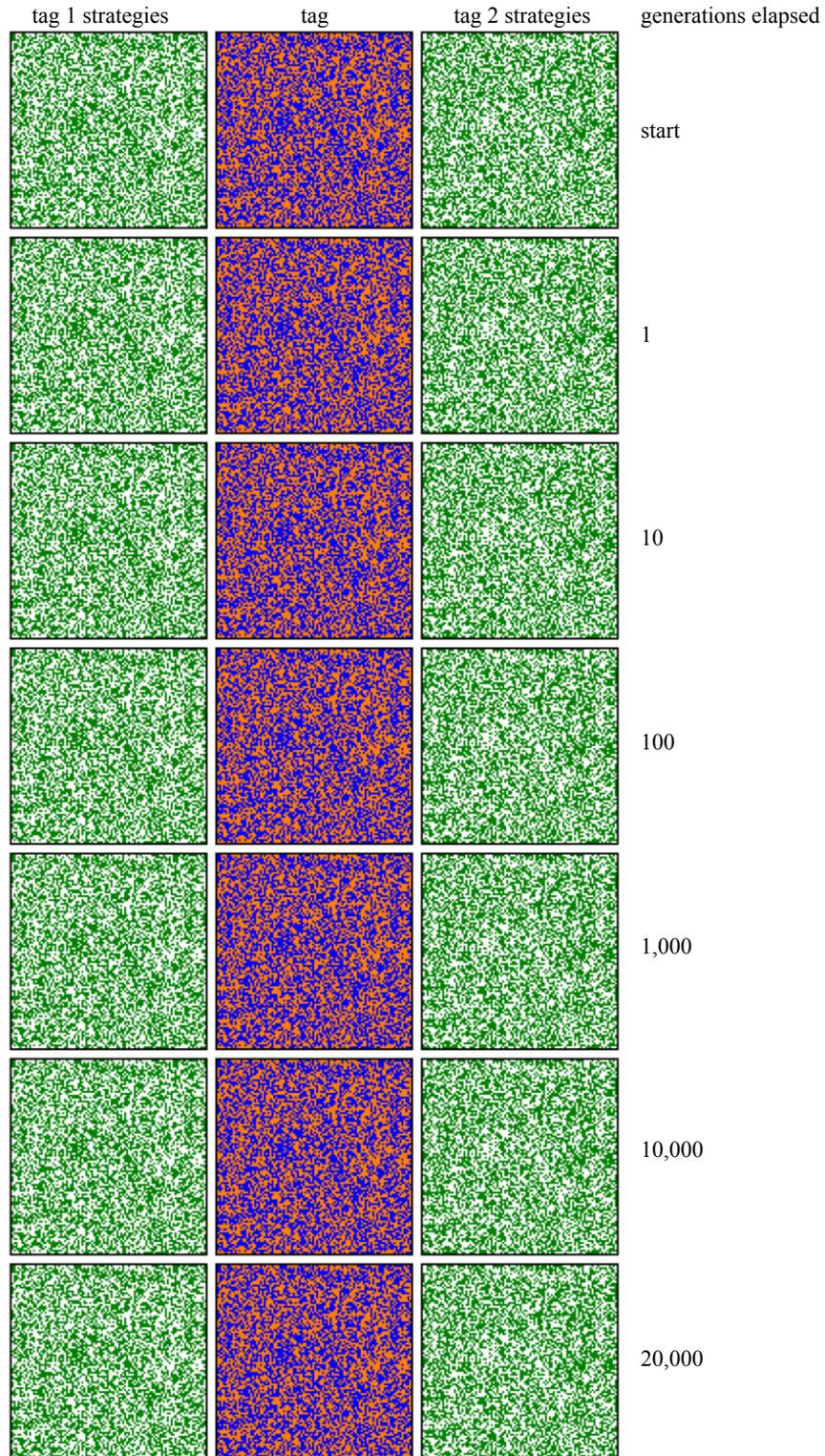
**Fig. A.4.** Fine-grain spatial partitioning by mixed-tag groups of extra-tag cooperators. For simplicity, this figure considers the one-parameter payoff matrix discussed in the Introduction (main text). (a) When a  $C_j$  mutant appears in a patch of  $C_i$  it does not trigger an invasion because the  $C_j$  individual has the same average payoff as the neighbouring  $C_i$  individuals ( $p_{C_j} = p_{C_i} = 1$ ). (d) Likewise,  $D_j$  does not invade  $D_i$  ( $p_{D_j} = p_{D_i} = 0$ ). (b) In the case of an  $I_j$  mutant arising in a patch of  $I_i$ , not only does  $I_j$  not invade, it is outcompeted by its  $I_i$  neighbours ( $p_{I_j} = 0$ ;  $p_{I_i} = 0.75$ ). (c) Conversely, when a  $E_j$  mutant appears in a patch of  $E_i$  it *does* trigger an invasion because it has a greater payoff than its neighbours ( $p_{E_j} = 1$ ;  $p_{E_i} = 0.25$ ). Thus, extra-tag cooperators of different tags are mutually invisable at the neighbourhood scale and readily form fine-grained mixed-tag coalitions. This is the only pair of tag-by-strategy types for which this is possible.

Figures A.5-A.13 show snapshots of example simulations of the main outcomes of spatial interactions between pairs of tag-by-strategy types, as described in Figs. A.1-A.4:

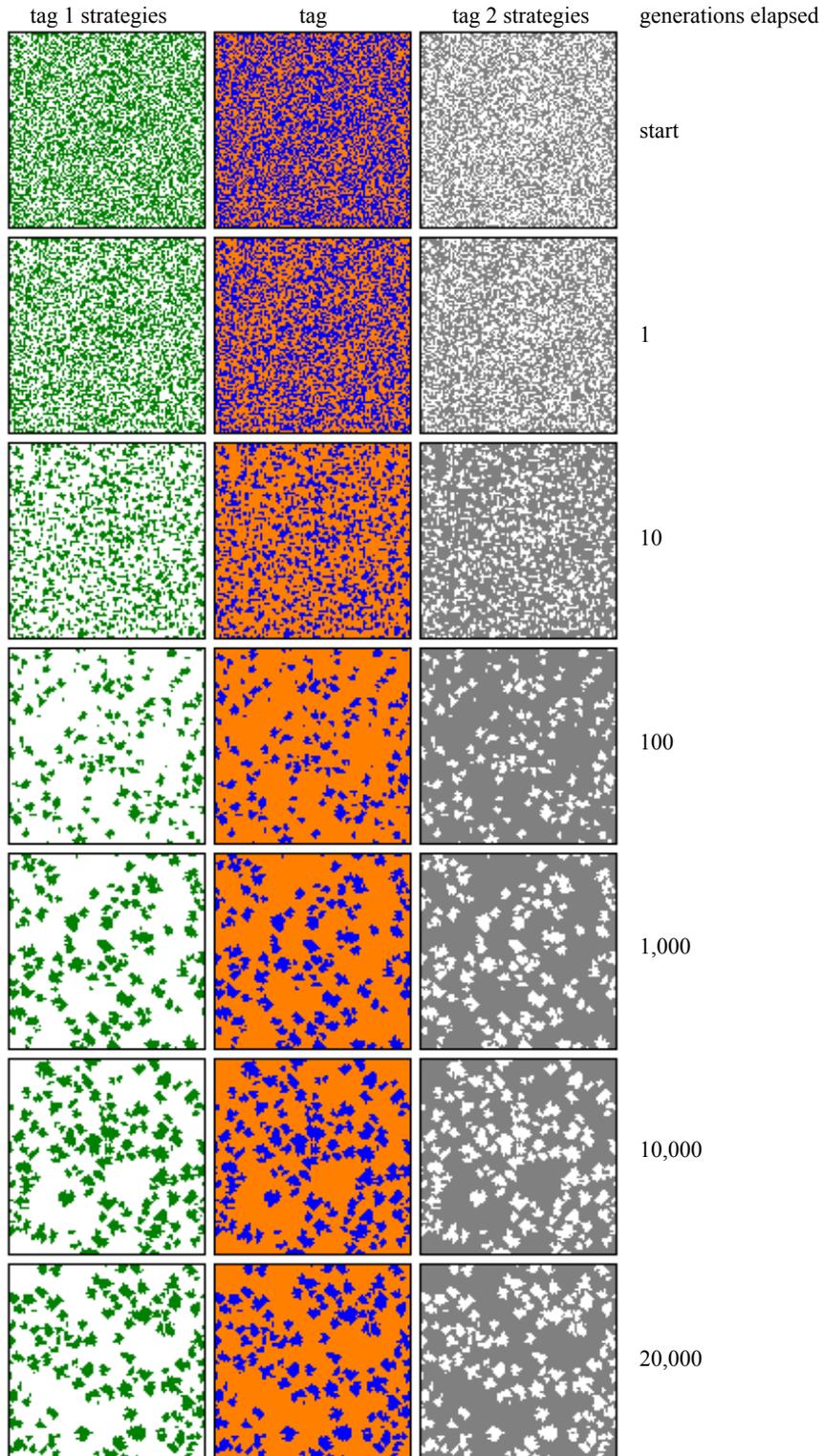
- Figure A.5 gives an example of a standoff that occurs when all interactions result in mutual cooperation, in this case  $C_i$  versus  $C_j$ . Note that  $C_i$  versus  $I_i$  is an equivalent situation.

- Figures A.6 and A.7 give examples of the situation where one tag-by-strategy type always cooperates (here,  $C_i$ ) and the other always defects (here,  $D_j$ ). As shown by [35], if the cost-benefit ratio of mutual cooperation,  $r$ , is less than approximately 0.07, the situational cooperator can persist by forming complex clusters (Fig. A.6); however, for higher values of  $r$  the situational defector takes over the population (Fig. A.7). Note that  $C_i$  versus  $D_i$ ,  $I_i$  versus  $E_i$ ,  $I_i$  versus  $D_i$ , and  $C_i$  versus  $E_i$  are all equivalent situations.
- Figure A.8 gives an example of a standoff that occurs when all the interactions result in mutual defection, in this case  $D_i$  versus  $D_j$ . Note that  $E_i$  versus  $D_i$  is an equivalent situation.
- Figure A.9 gives an example of the situation where one tag-by-strategy type overtakes the other regardless of the value of the cost-benefit ratio of mutual cooperation ( $r$ ), in this case  $C_i$  versus  $I_j$  (the latter being the winner). Note that  $I_i$  versus  $E_j$ ,  $I_i$  versus  $D_j$ , and  $D_i$  versus  $E_j$  are very similar situations.
- Figure A.10 gives an example of the scenario of  $C_i$  versus  $E_j$ . In this case,  $C$  overtakes clusters of  $E$  until all  $E$  individuals are isolated; after that, a standoff occurs due to all further interactions being mutually cooperative.
- Figures A.11 and A.12 give examples of the scenario of  $I_i$  versus  $I_j$ . The two types are equally matched, on average, so this often results in a quasi-standoff between them (Fig. A.11); however, if one of the types gains the ‘upper hand’ and surrounds the other, then along the borders, the average payoff of the *surrounding* type will be greater than that of the *surrounded* type (e.g., Fig. A.4b), leading to slow extinction of the latter (Fig. A.12).
- Figure A.13 gives an example of the situation of  $E_i$  versus  $E_j$ . Due to negative frequency dependence, the two types are mutually miscible at the local scale. This results in the formation of mixed-tag aggregations, which can be thought of as a ‘meta-strategy’.

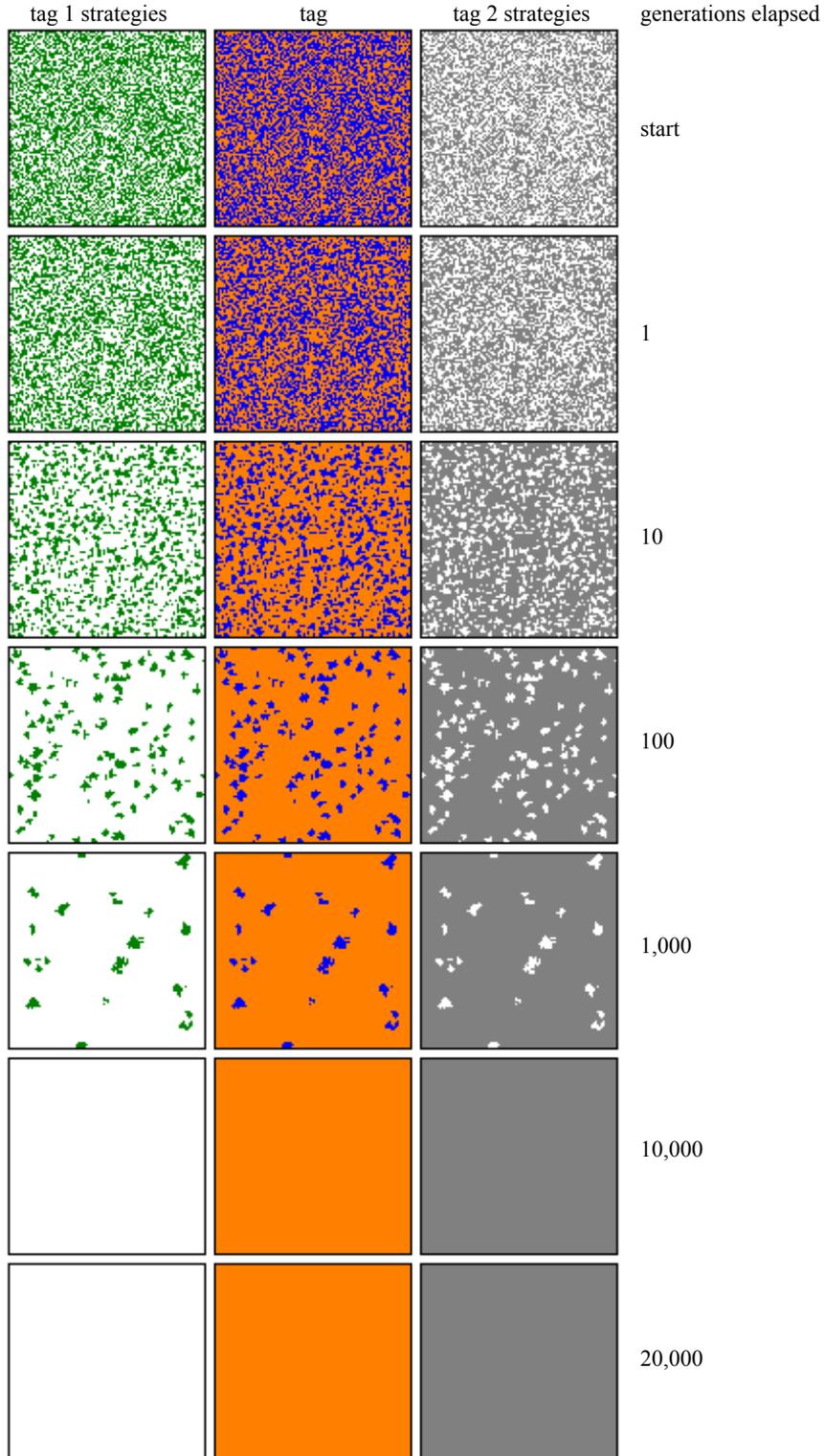
Interestingly, the domain-growth processes seen in Fig. A.11, for example, have been studied in the context of the link between statistical physics and game theory (see Szabó et al. 2000, Phys. Rev. E 62, 1095-1103, Szabó et al. 2010, Phys. Rev. E 82, 026110, and references therein).



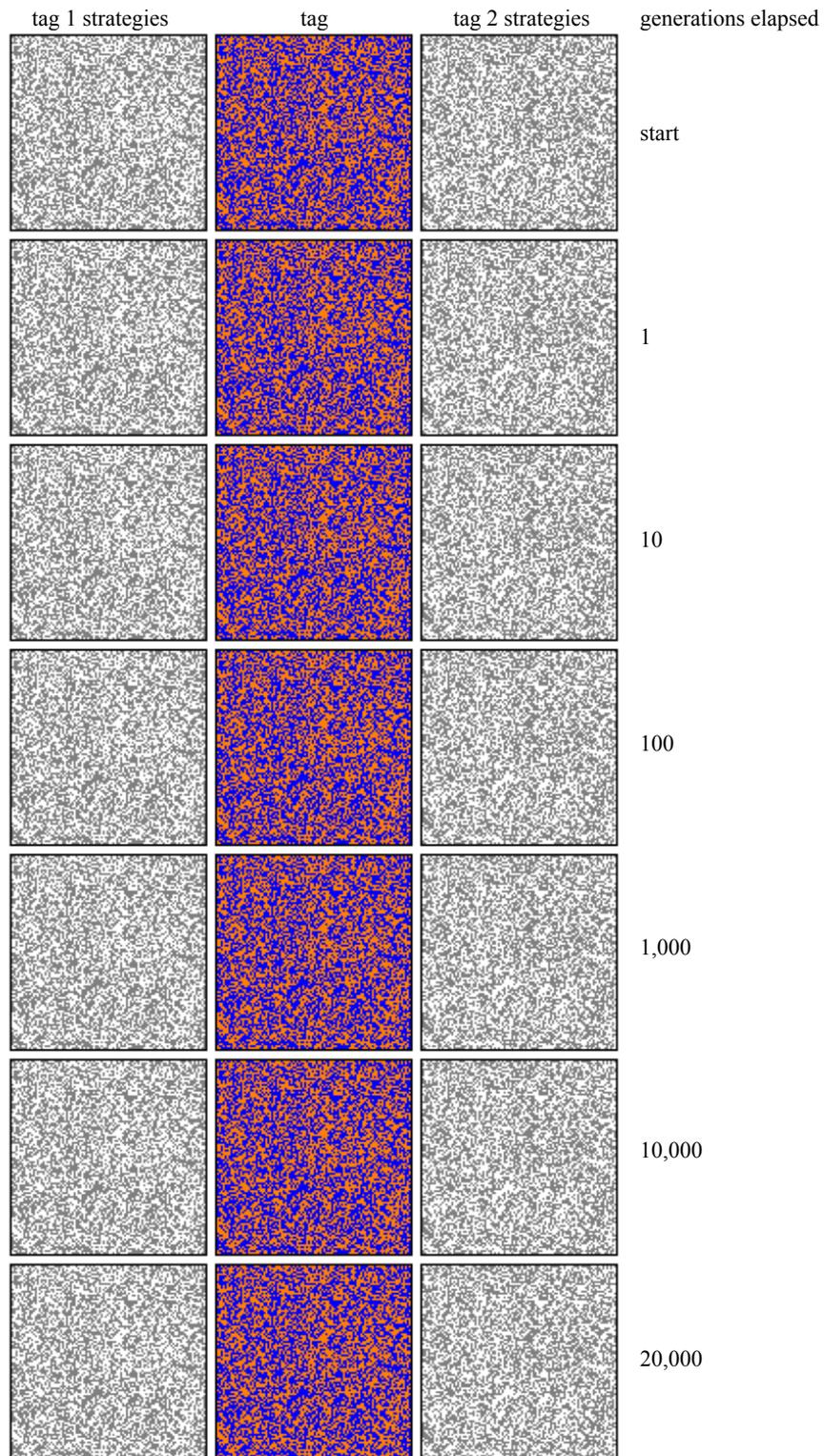
**Fig. A.5.** Snapshots of spatial lattice model for  $C_i$  versus  $C_j$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $C = \text{green}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $C = \text{green}$ ). This pair of tag-by-strategy types results in a standoff for all  $r$ .



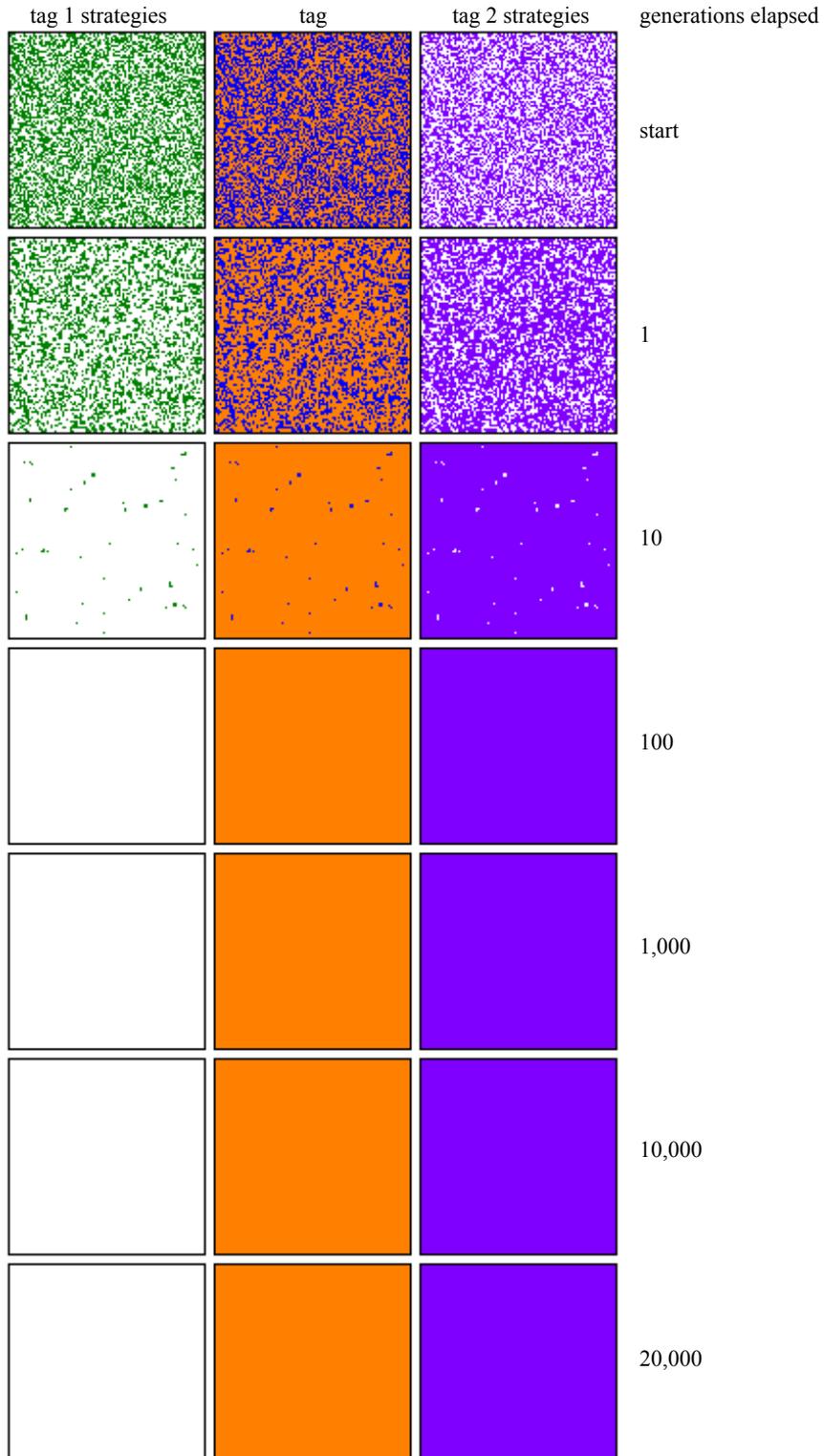
**Fig. A.6.** Snapshots of spatial lattice model for  $C_i$  versus  $D_j$  ( $\mu = 0$ ,  $r = 0.06$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $C = \text{green}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $D = \text{grey}$ ). This pair of tag-by-strategy types results in coexistence when  $r$  is less than approximately 0.07.



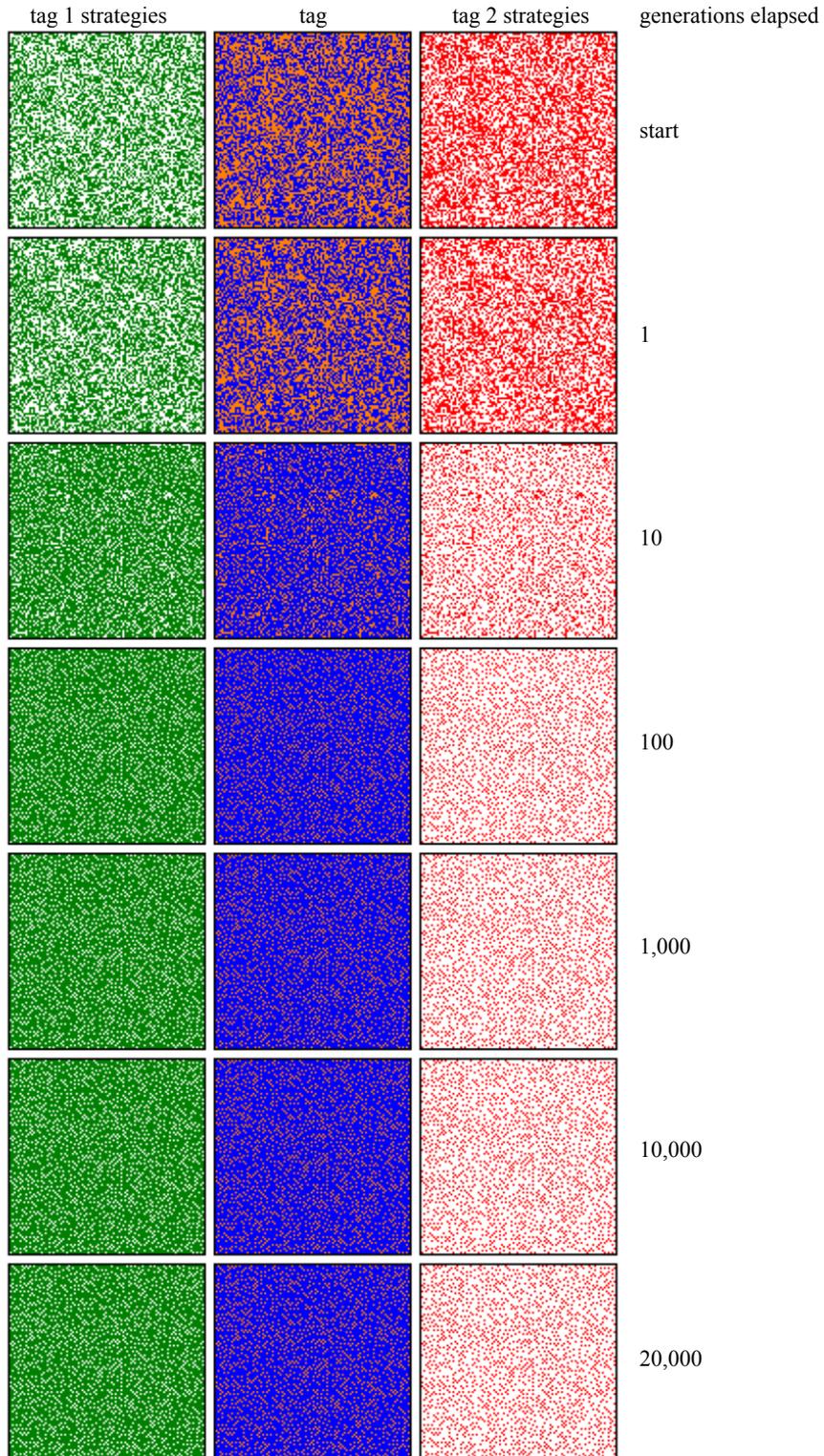
**Fig. A.7.** Snapshots of spatial lattice model for  $C_i$  versus  $D_j$  ( $\mu = 0$ ,  $r = 0.08$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $C = \text{green}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $D = \text{grey}$ ). This pair of tag-by-strategy types results in  $D$ 's dominance when  $r$  is greater than approximately 0.07.



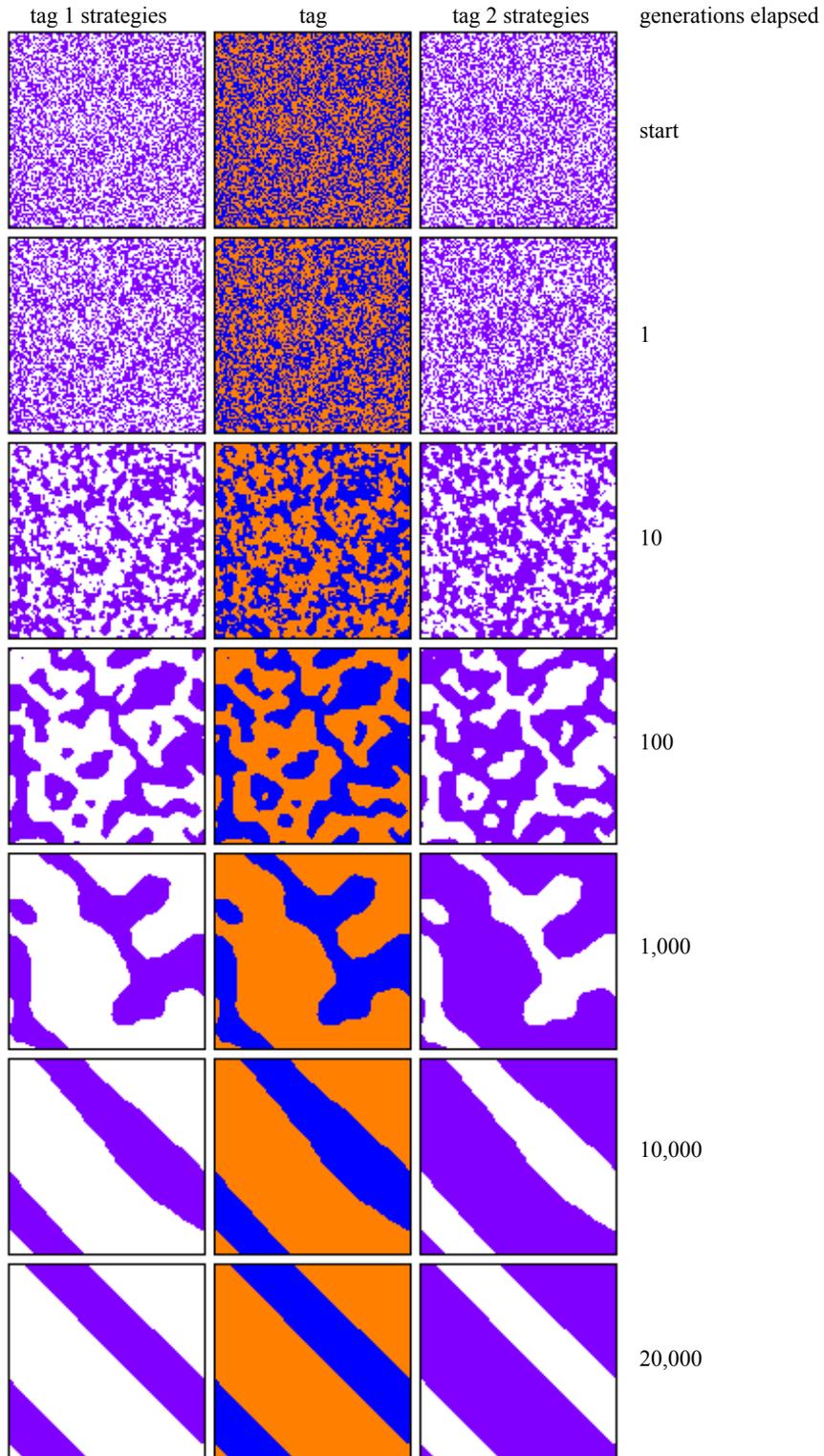
**Fig. A.8.** Snapshots of spatial lattice model for  $D_i$  versus  $D_j$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $D = \text{grey}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $D = \text{grey}$ ). This pair of tag-by-strategy types results in a standoff for all  $r$ .



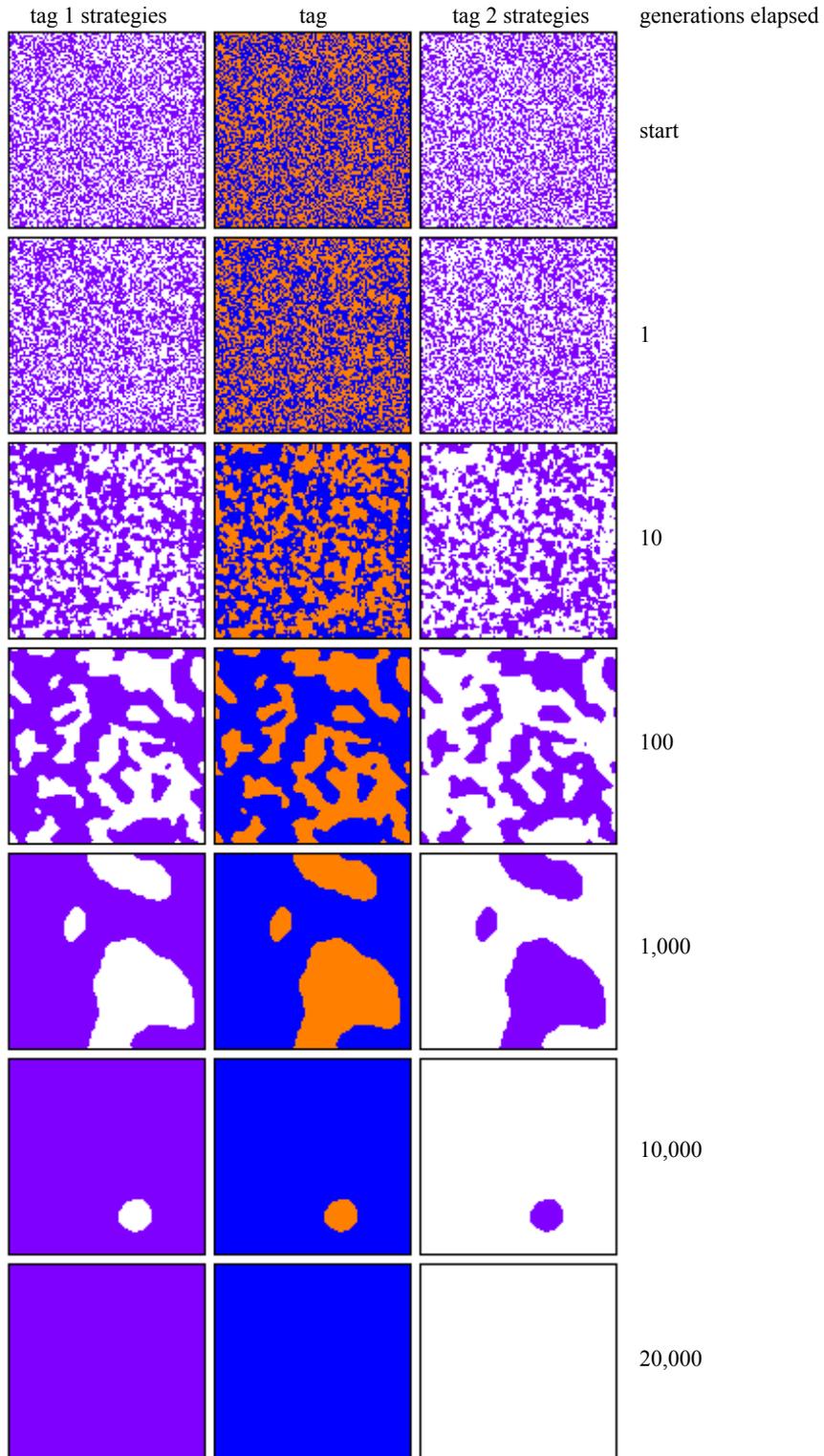
**Fig. A.9.** Snapshots of spatial lattice model for  $C_i$  versus  $I_j$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $C = \text{green}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $I = \text{purple}$ ). This pair of tag-by-strategy types results in  $I$ 's dominance for all  $r$ .



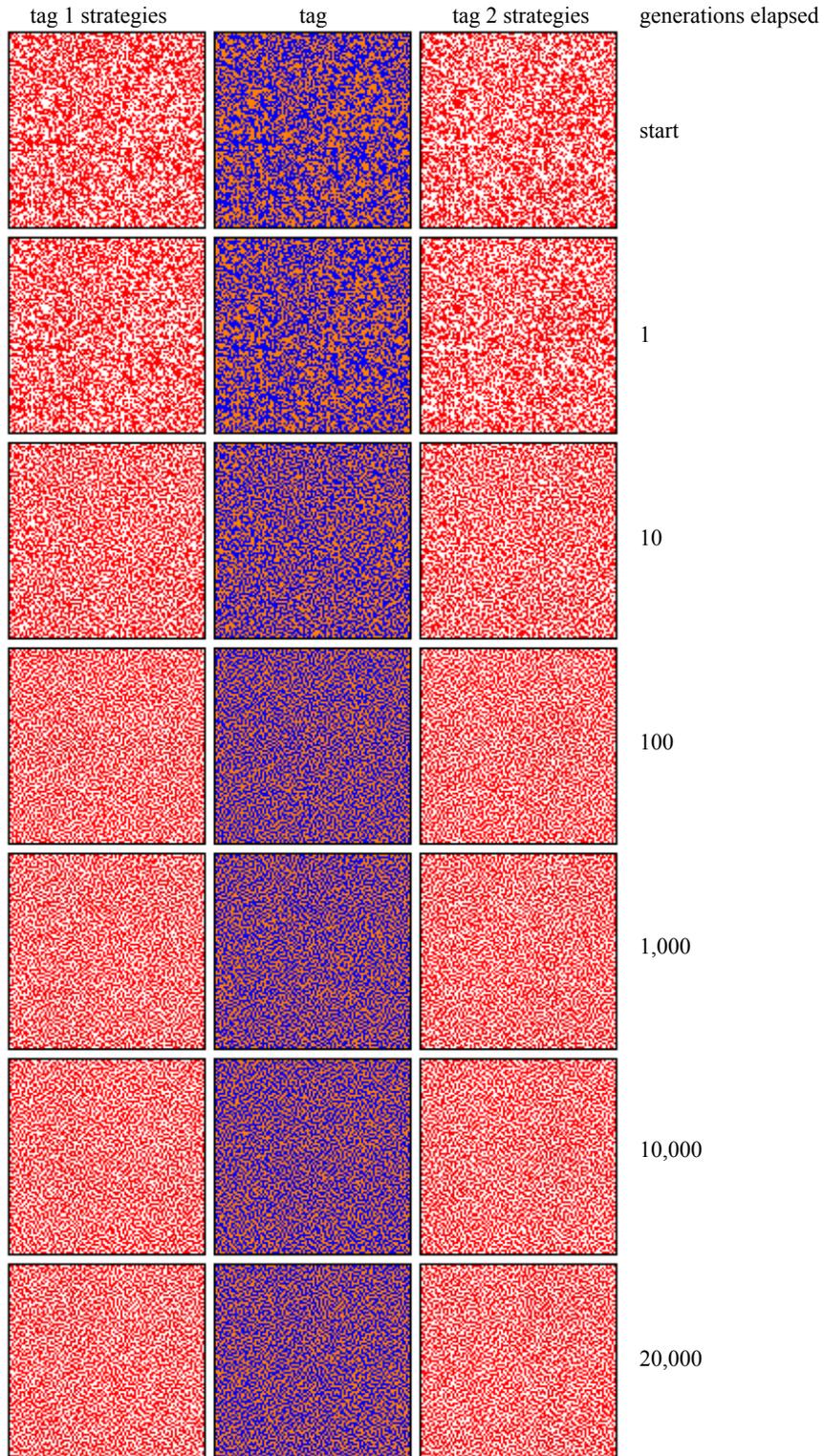
**Fig. A.10.** Snapshots of spatial lattice model for  $C_i$  versus  $E_j$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $C = \text{green}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $E = \text{red}$ ). This pair of tag-by-strategy types results isolation of  $E$ , and then a standoff, for all  $r$ .



**Fig. A.11.** Snapshots of spatial lattice model for  $I_i$  versus  $I_j$  ( $\mu = 0$ ,  $r = 0.32$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $I =$  purple). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals (again,  $I =$  purple). This pair of tag-by-strategy types sometimes results in a quasi-standoff, depending on the spatial configuration.



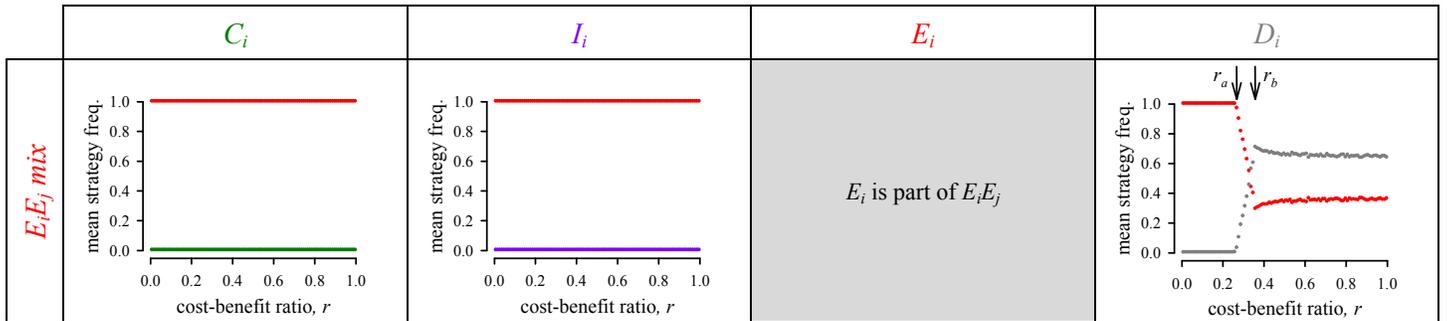
**Fig. A.12.** Snapshots of spatial lattice model for  $I_i$  versus  $I_j$  ( $\mu = 0$ ,  $r = 0.31$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $I =$  purple). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals (again,  $I =$  purple). This pair of tag-by-strategy types sometimes results in extinction of one of the tags, depending on the spatial configuration.



**Fig. A.13.** Snapshots of spatial lattice model for  $E_i$  versus  $E_j$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $E = \text{red}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals (again,  $E = \text{red}$ ). This pair of tag-by-strategy types results in mutual miscibility, independently of  $r$ .

## Appendix B: Outcomes involving $E_iE_j$ aggregations in spatially structured populations with no mutation

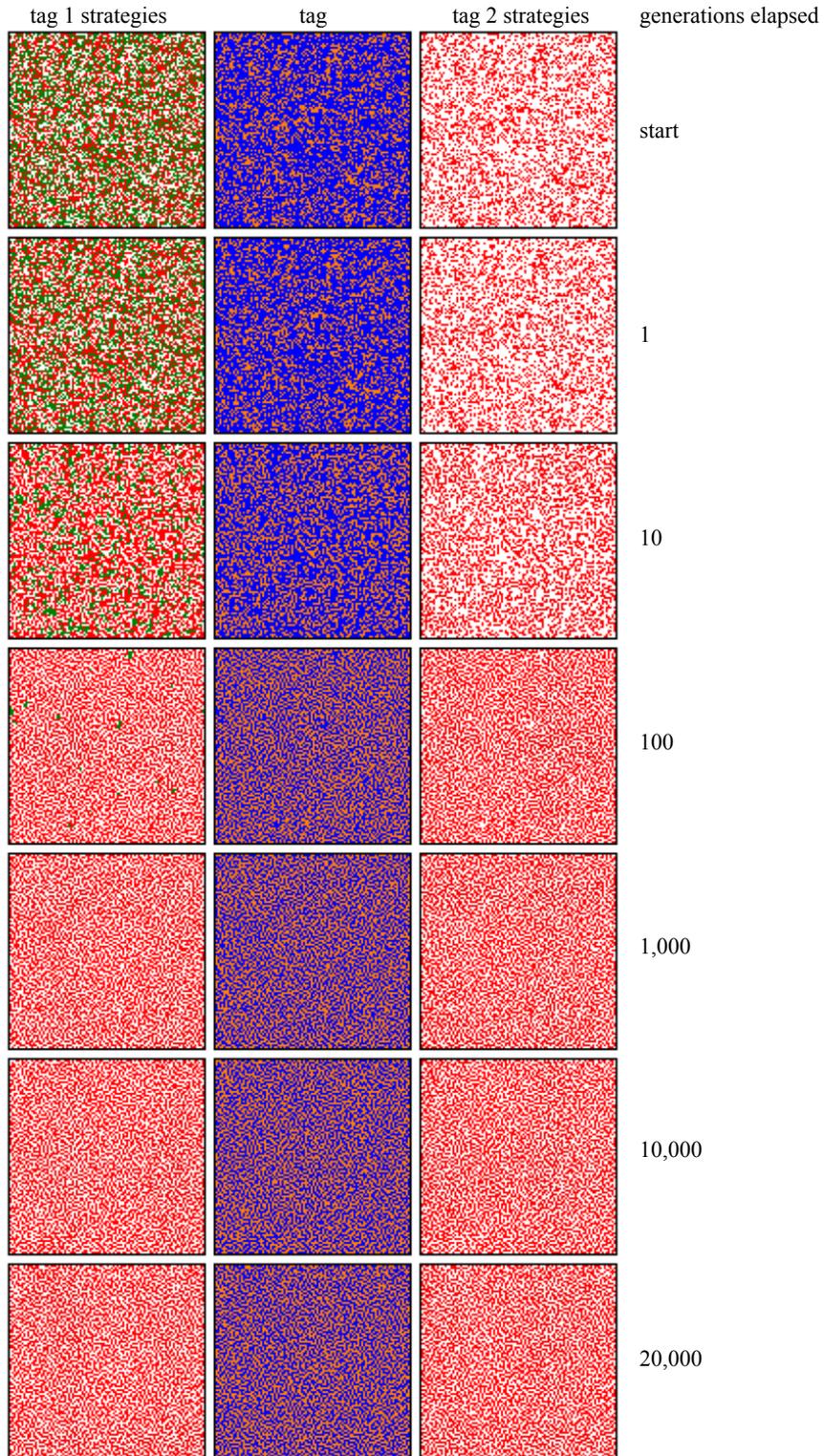
Differently tagged  $E$ -strategists are mutually miscible and readily form fine-grained mixed-tag aggregations that can be thought of as ‘meta-strategies’ (see §3.1 in the main text and Appendix A). Therefore, for two-tag systems, it is instructive to ascertain the outcome of spatial interactions between these mixed-tag aggregations ( $E_iE_j$ ) and the other strategies ( $C$ ,  $I$ , and  $D$ ) in the absence of mutation. Figure B.1 shows that  $E_iE_j$  beats  $C_i$  and  $I_i$  for all values of the cost-benefit ratio of mutual cooperation,  $r$ .  $E_iE_j$  can eliminate  $D_i$  when  $r$  is relatively low ( $r < r_a \approx 0.27$ ), but for increasing  $r$ ,  $D_i$  is able to first persist with  $E_iE_j$  ( $r_a < r < r_b \approx 0.36$ ), and then even purge  $E_j$  from the mixed-tag aggregation ( $r > r_b$ ) leaving a  $E_i$ - $D_i$  standoff (which, if other tag-by-strategy types are also present, makes  $E_i$  very vulnerable to exploitation; see Appendix A and Table 1b, main text).



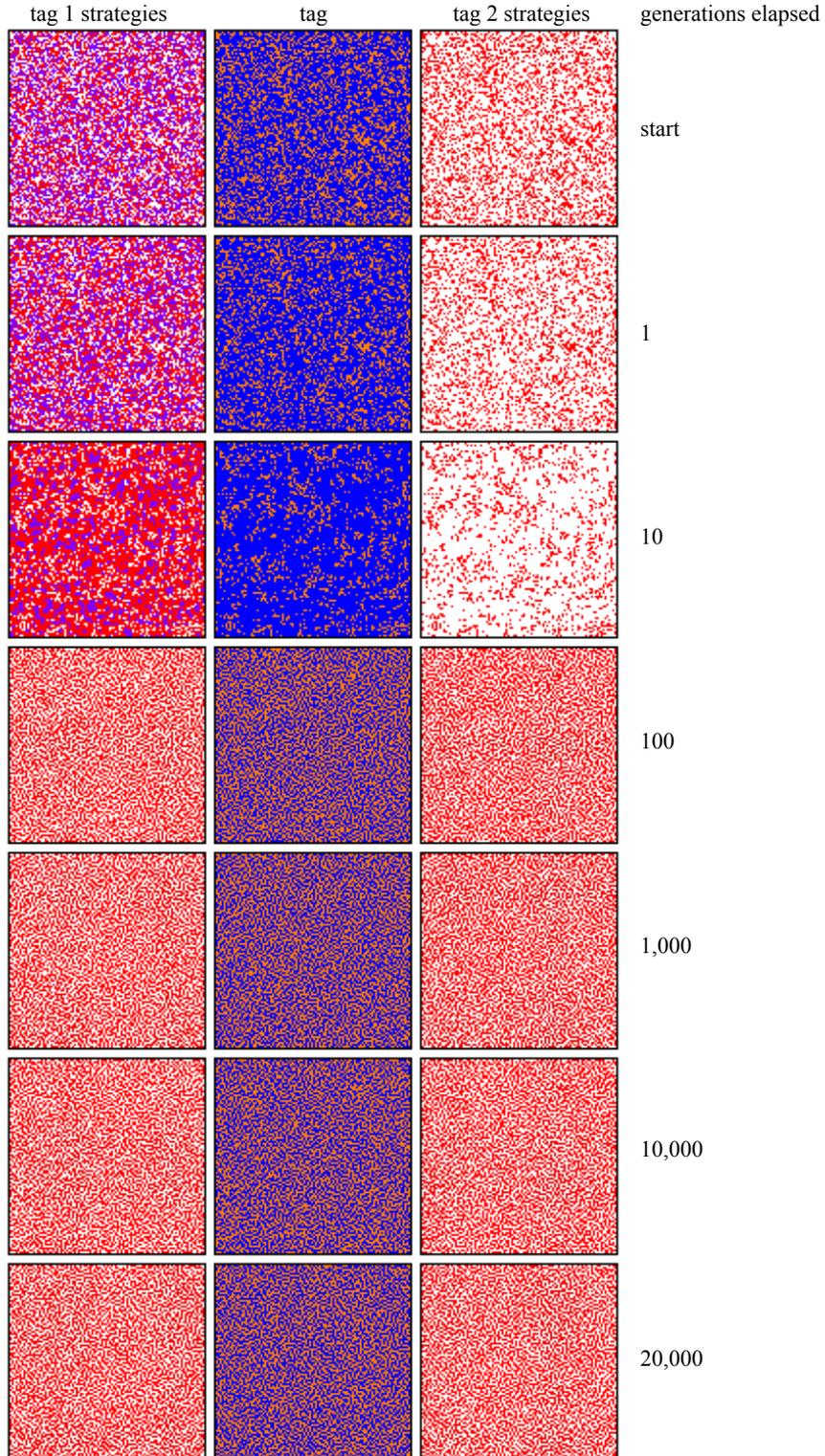
**Fig. B.1.** Spatial invasion outcomes when only three strategy-by-tag types are present (i.e., a mixed-tag aggregation,  $E_iE_j$ , and one other type, given by the column header) and there is no mutation. This figure demonstrates through simulations the results presented in Table 1c and §3.1 in the main text. Each panel shows the average strategy frequency between 15- and 20-thousand generations of the spatial model, for populations of 10,000 individuals, and cost benefit ratios of  $0 < r < 1$  in increments of 0.01. Initially there was a random arrangement of strategies. Line colours correspond to the colours of the row and column headers, with  $E_i$  and  $E_j$  shown in aggregate.  $E_i$  is part of  $E_iE_j$  and is therefore subsumed by it.

Figures B.2-B.6 show snapshots of example simulations of the main outcomes of spatial interactions between tag-by-strategy types, as described in Fig. B.1:

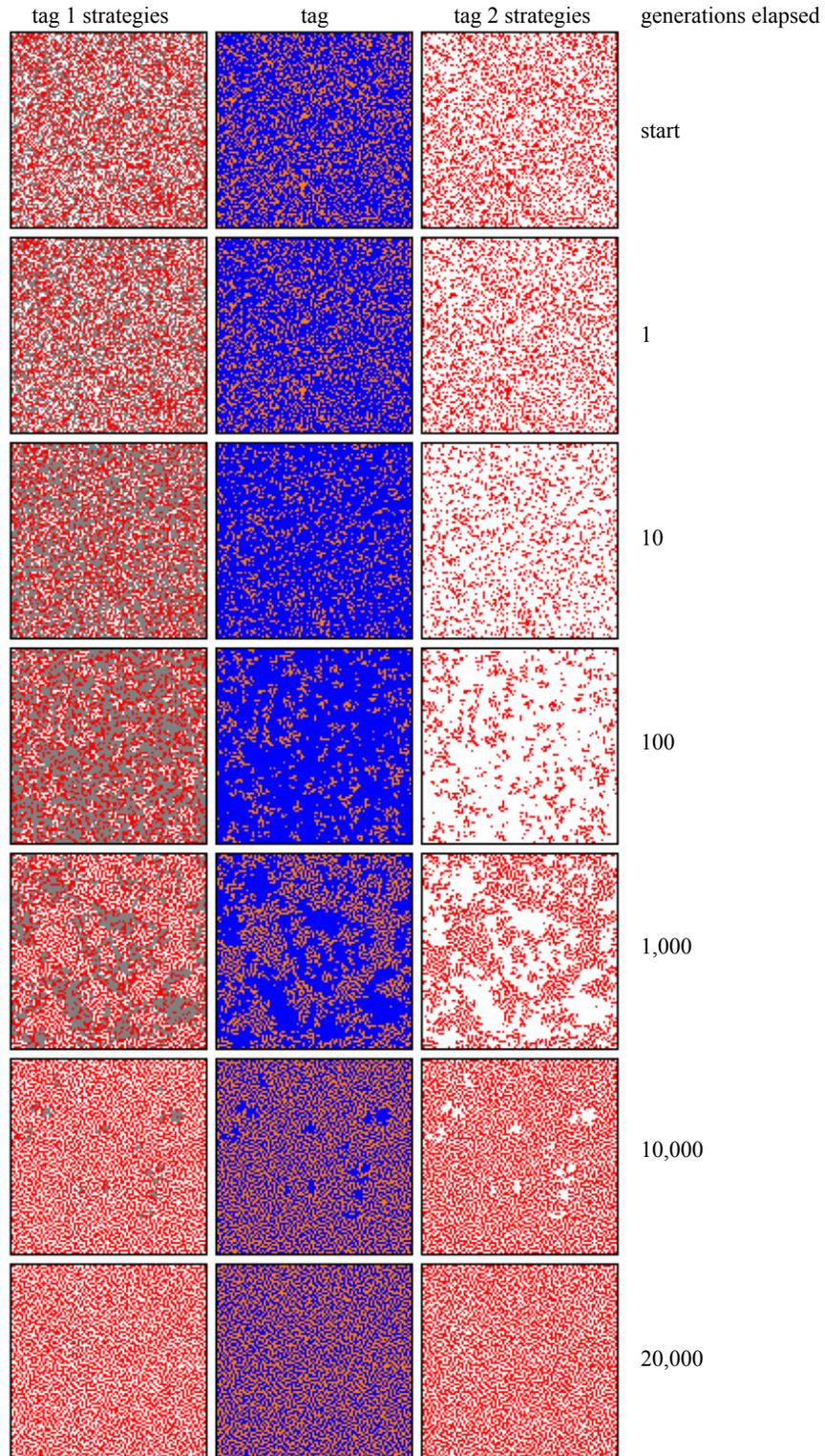
- Figure B.2 gives an example of a mixed-tag aggregation of  $E_iE_j$  overtaking  $C_i$ .
- Figure B.3 gives an example of a mixed-tag aggregation of  $E_iE_j$  overtaking  $I_i$ .
- Figure B.4 gives an example of a mixed-tag aggregation of  $E_iE_j$  overtaking  $D_i$  at low  $r$ .
- Figure B.5 gives an example of a mixed-tag aggregation of  $E_iE_j$  coexisting with  $D_i$  at intermediate  $r$ .
- Figure B.6 gives an example of  $D_i$  purging  $E_j$  from a mixed-tag aggregation of  $E_iE_j$  at high  $r$ , followed by a standoff.



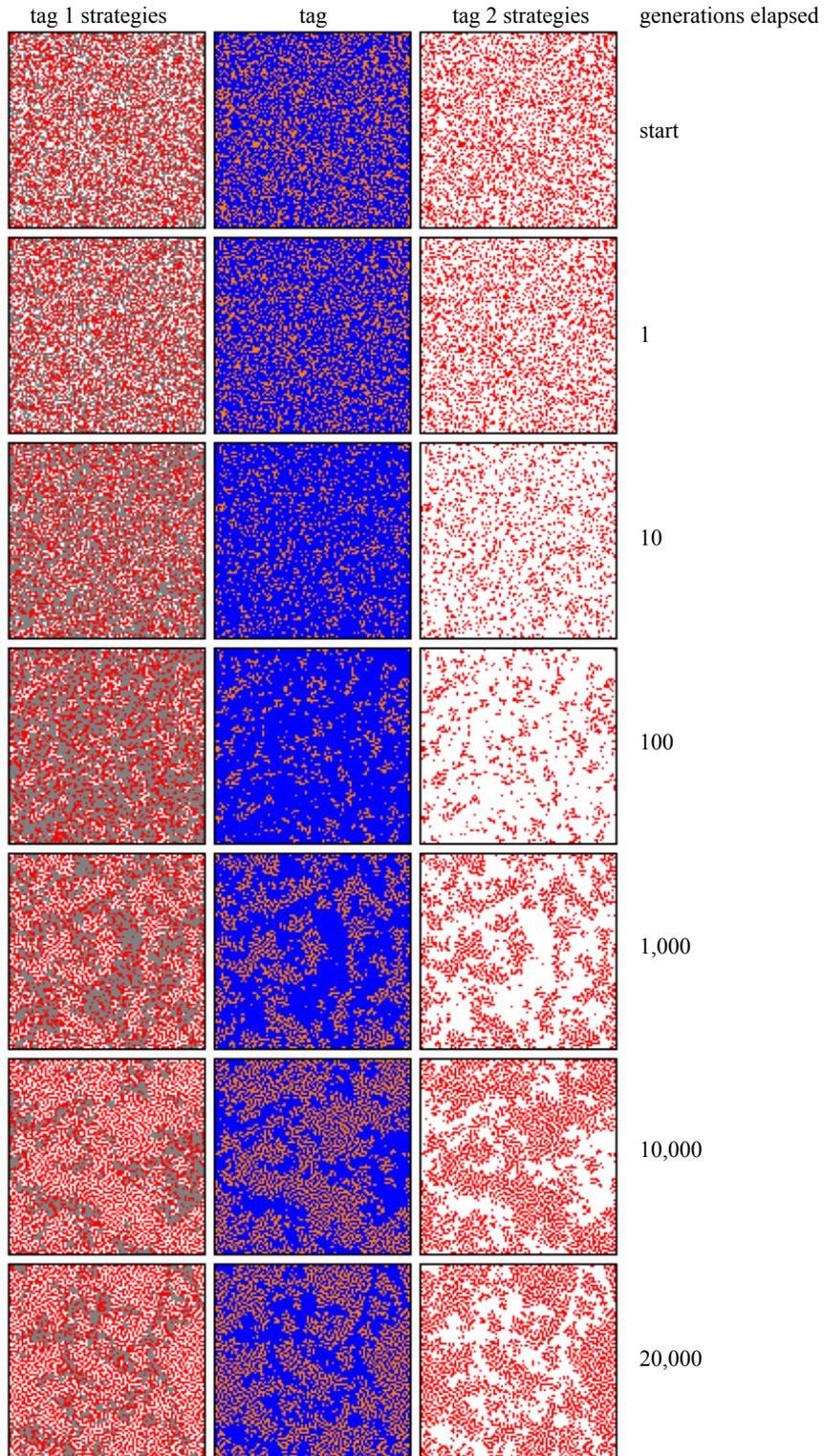
**Fig. B.2.** Snapshots of spatial lattice model for  $E_i E_j$  versus  $C_i$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $E = \text{red}$ ,  $C = \text{green}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals (again,  $E = \text{red}$ ).  $E_i E_j$  overtakes  $C_i$ , independently of  $r$ .



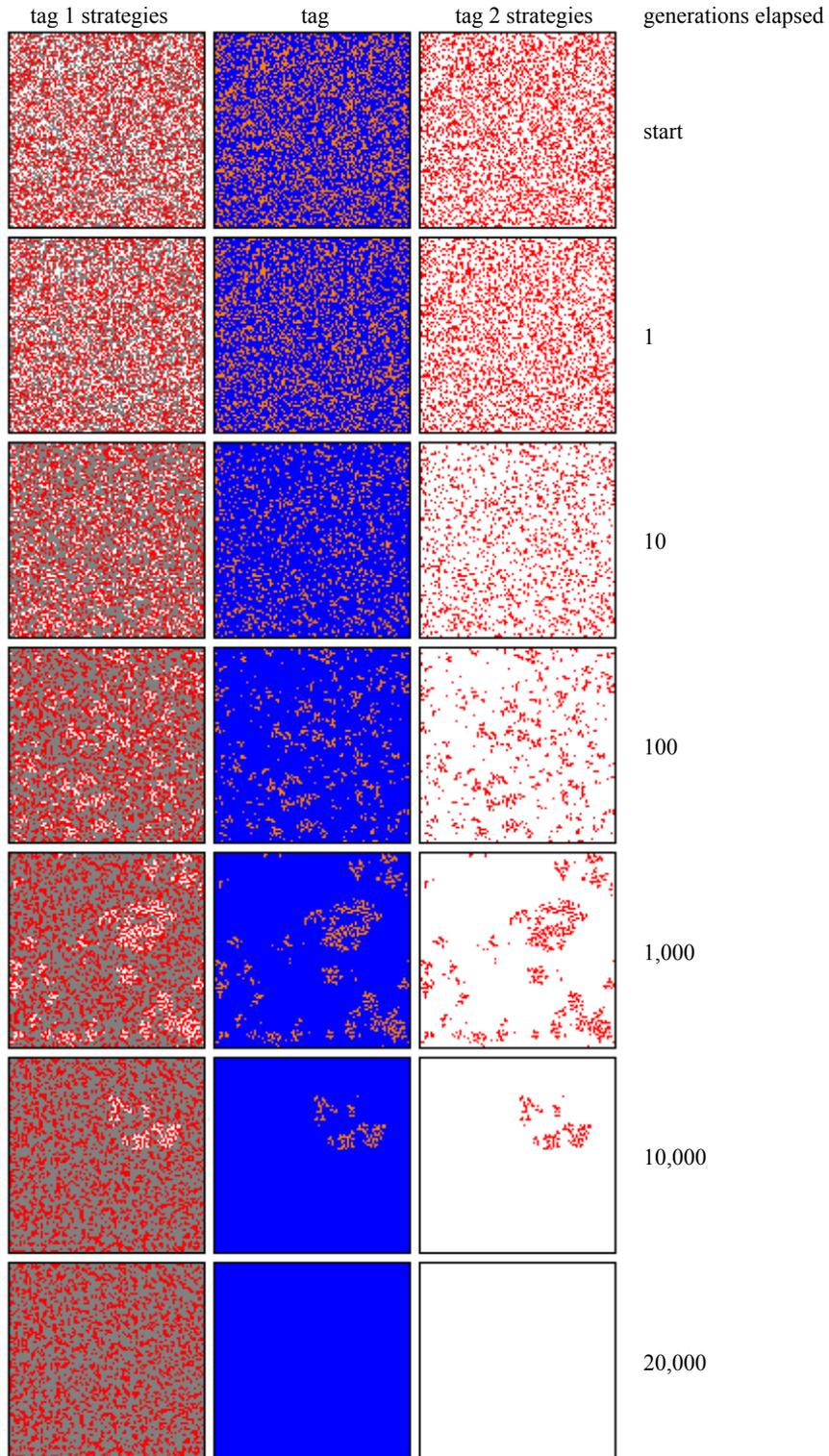
**Fig. B.3.** Snapshots of spatial lattice model for  $E_i E_j$  versus  $I_i$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $E = \text{red}$ ,  $I = \text{purple}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals (again,  $E = \text{red}$ ).  $E_i E_j$  overtakes  $I_i$ , independently of  $r$ .



**Fig. B.4.** Snapshots of spatial lattice model for  $E_i E_j$  versus  $D_i$  ( $\mu = 0$ ,  $r = 0.26$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $E = \text{red}$ ,  $D = \text{grey}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $E = \text{red}$ ).  $E_i E_j$  overtakes  $D_i$  for low values of  $r$ .



**Fig. B.5.** Snapshots of spatial lattice model for  $E_i E_j$  versus  $D_i$  ( $\mu = 0$ ,  $r = 0.29$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $E = \text{red}$ ,  $D = \text{grey}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $E = \text{red}$ ).  $E_i E_j$  and  $D_i$  coexist for intermediate values of  $r$ .

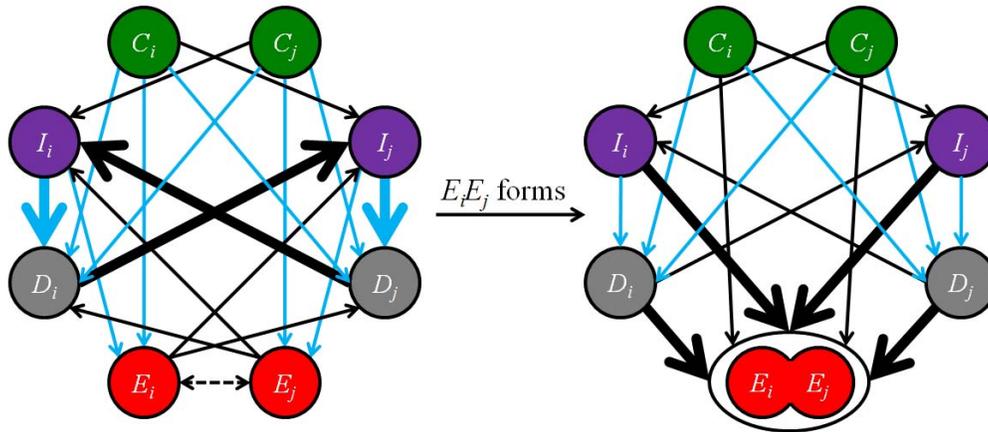


**Fig. B.6.** Snapshots of spatial lattice model for  $E_i E_j$  versus  $D_i$  ( $\mu = 0$ ,  $r = 0.38$ ). The lattice is shown as three corresponding panels for seven generations (start-20,000). The middle panel shows the spatial positions of Tag  $i = 1$  (blue) and Tag  $j = 2$  (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $E = \text{red}$ ,  $D = \text{grey}$ ). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals ( $E = \text{red}$ ). At high values of  $r$ ,  $D_i$  purges  $E_j$  from the mixed-tag aggregation  $E_i E_j$ , leading to a standoff between  $D_i$  and  $E_i$ , caused by the fact that all the interactions between  $D_i$  and  $E_i$  result in mutual defection.

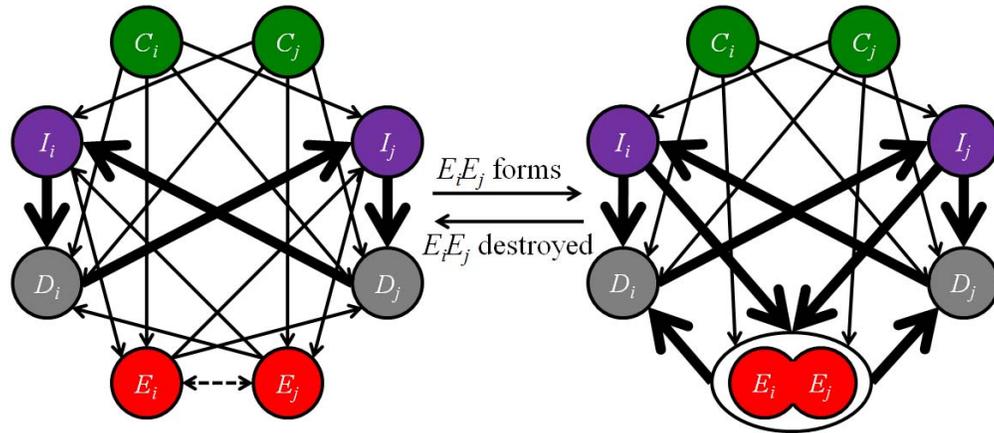
### Appendix C: Graphical interpretation of outcomes of the spatial, two-tag Prisoner's Dilemma game

The results from the main two-tag spatial model (Figs. 1-4 in the main text) can be interpreted in terms of the outcomes of the pair-wise interactions between tag-by-strategy types in the spatial arena (Appendices A, B):

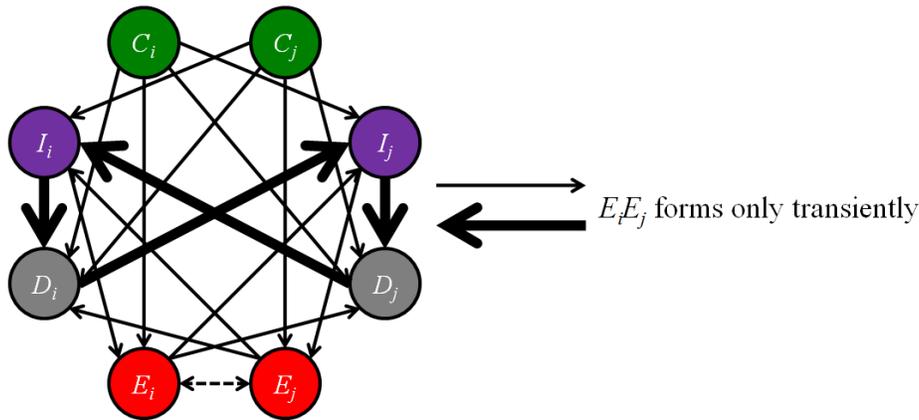
- When the cost-benefit ratio of mutual cooperation,  $r = c/(b - c)$ , is low (less than approximately 0.16), mixed-tag aggregations of  $E$ -strategists take over the lattice (Fig. C.1; also see Fig. 4 in the main text and Movies D.1 and D.9 for example dynamics).
- When  $r$  is of intermediate value (between approximately 0.16 and 0.41),  $I$ -,  $E$ -, and  $D$ -strategists coexist over evolutionary time (Fig. C.2; also see Fig. 3 in the main text and Movies D.2 and D.10 for example dynamics).
- When  $r$  is of a higher value (between approximately 0.41 and 0.62), only  $I$ - and  $D$ -strategists coexist (Fig. C.3; also see Fig. 2 in the main text and Movies D.3 and D.11 for example dynamics).
- When  $r$  is very high (greater than approximately 0.62),  $D$ -strategists dominate the population (Fig. C.4; also see Movies D.4 and D.12 for example dynamics).



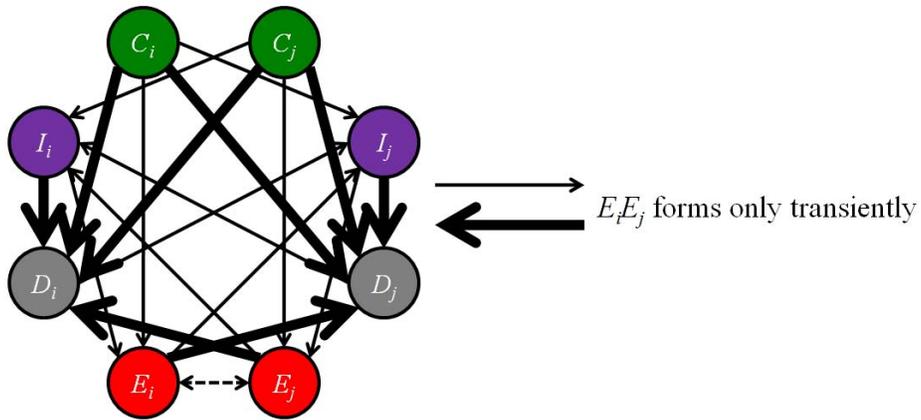
**Fig. C.1.** Two-tag Prisoner's Dilemma for  $r < r_1$  ( $r_1 \approx 0.16$ ). Arrows point from competitive subordinate to dominant in pair-wise spatial interactions (see Figs. A.1, A.2). Thick arrows denote the dominant interactions. The dotted arrow indicates mutual miscibility, allowing mixed-tag aggregations to form. Blue arrows indicate that the subordinate tag-by-strategy can actually persist during pair-wise spatial interactions by forming compact clusters, provided that the cost-benefit ratio of mutual cooperation,  $r$ , is very low (less than approximately 0.07). 'Standoffs' are not included. The left side indicates what happens before mixed-tag aggregations of  $E$ -strategists ( $E_i E_j$ ) form (dynamics initially dominated by non-transitive spatial invasibility of the form  $I_i \rightarrow D_i \rightarrow I_j \rightarrow D_j \rightarrow I_i$ ). The right side indicates what happens after  $E_i E_j$  aggregations form ( $E_i E_j$  eventually dominates). See Fig. 4 in the main text and Movies D.1 and D.9 for example dynamics.



**Fig. C.2.** Two-tag Prisoner's Dilemma for  $r_1 < r < r_2$  ( $r_1 \approx 0.16$ ,  $r_2 \approx 0.41$ ). Arrows point from competitive subordinate to dominant in pair-wise spatial interactions (see Figs. A.1, A.2). Thick arrows denote the dominant interactions. The dotted arrow indicates mutual miscibility, allowing mixed-tag aggregations to form. 'Standoffs' are not included. The left side indicates what happens before mixed-tag aggregations of  $E$ -strategists ( $E_i E_j$ ) form (dynamics initially dominated by non-transitive spatial invasibility of the form  $I_i \rightarrow D_i \rightarrow I_j \rightarrow D_j \rightarrow I_i$ ). The right side indicates what happens *after*  $E_i E_j$  aggregations form (at first,  $E_i E_j$  grows at the expense of  $I_i$  and  $I_j$ , but it is simultaneously overtaken by  $D_i$  and  $D_j$ ; eventually, this latter interaction destroys the mixed-tag aggregation leading to a long-term cycle involving  $I$ -,  $E$ -, and  $D$ -strategists). See Fig. 3 in the main text and Movies D.2 and D.10 for example dynamics.



**Fig. C.3.** Two-tag Prisoner's Dilemma for  $r_2 < r < r_3$  ( $r_2 \approx 0.41$ ,  $r_3 \approx 0.62$ ). Arrows point from competitive subordinate to dominant in pair-wise spatial interactions (see Figs. A.1, A.2). Thick arrows denote the dominant interactions. The dotted arrow indicates mutual miscibility, potentially allowing mixed-tag aggregations to form. 'Standoffs' are not included. The left side indicates what happens before mixed-tag aggregations of  $E$ -strategists ( $E_i E_j$ ) form (dynamics dominated by non-transitive spatial invasibility of the form  $I_i \rightarrow D_i \rightarrow I_j \rightarrow D_j \rightarrow I_i$ ). Unlike for lower values of  $r$  (Figs. C.1, C.2), here  $E_i E_j$  aggregations are rapidly destroyed by  $D$ -strategists, and do not ever come to prominence within the population. See Fig. 2 in the main text and Movie D.3 for example dynamics.



**Fig. C.4.** Two-tag Prisoner's Dilemma for  $r > r_3$  ( $r_3 \approx 0.62$ ). Arrows point from competitive subordinate to dominant in pair-wise spatial interactions (see Figs. A.1, A.2). Thick arrows denote the dominant interactions. The dotted arrow indicates mutual miscibility, potentially allowing mixed-tag aggregations to form. 'Standoffs' are not included. The left side indicates what happens before mixed-tag aggregations of  $E$ -strategists ( $E_i E_j$ ) form ( $D_i$  and  $D_j$  take over). Unlike for lower values of  $r$  (Figs. C.1, C.2), here  $E_i E_j$  aggregations are rapidly destroyed by  $D$ -strategists, and do not ever come to prominence within the population. See Movie D.4 for example dynamics.

### Appendix D: Example model runs for the spatial, two-tag Prisoner's Dilemma game

There are a total of twelve hyperlinks to QuickTime movies for the parameter values and starting conditions shown in Table D.1; high-resolution versions of these movies are available by contacting the author. Movies D.1-D.4 give examples of the spatial dynamics of the two-tag Prisoner's Dilemma for the four regions delineated by  $r_1$ ,  $r_2$ , and  $r_3$  in Fig. 1 of the main text; Movies D.5-D.12 give further examples for other mutation rates and starting conditions, as discussed in the Appendices. Each movie shows the state of the 100 x 100 lattice every 10 model generations (i.e., every 100,000 interactions). The lattice is shown as three corresponding panels. The middle panel shows the spatial positions of Tag 1 (blue) and Tag 2 (orange). The left panel shows the spatial positions of the strategies of Tag 1 individuals ( $C$  = green,  $I$  = purple,  $E$  = red,  $D$  = grey). Similarly, the right panel shows the spatial positions of the strategies of Tag 2 individuals (again,  $C$  = green,  $I$  = purple,  $E$  = red,  $D$  = grey). For both the left and right panels, white regions are comprised of members of the opposite tag.

**Table D.1.** Parameter values and starting conditions for supplementary movies illustrating the different spatial tag- and strategy-dynamics discussed in the main text and the appendices. For the 'starting conditions' column, 'mixed' means that the tag and strategy of each cell was initially determined randomly and independently; 'ancestral' means that each cell was initially occupied by an unconditional defector of Tag 1 ( $D_1$ ). With one exception, Movie D.5, the movies show 20,000 model generations; the model run associated with Movie D.5 was truncated after it reached a monoculture, which is an evolutionary endpoint when there is no mutation. Note that these movies represent unique model runs and not the ones used to make the figures in the main text and appendices.

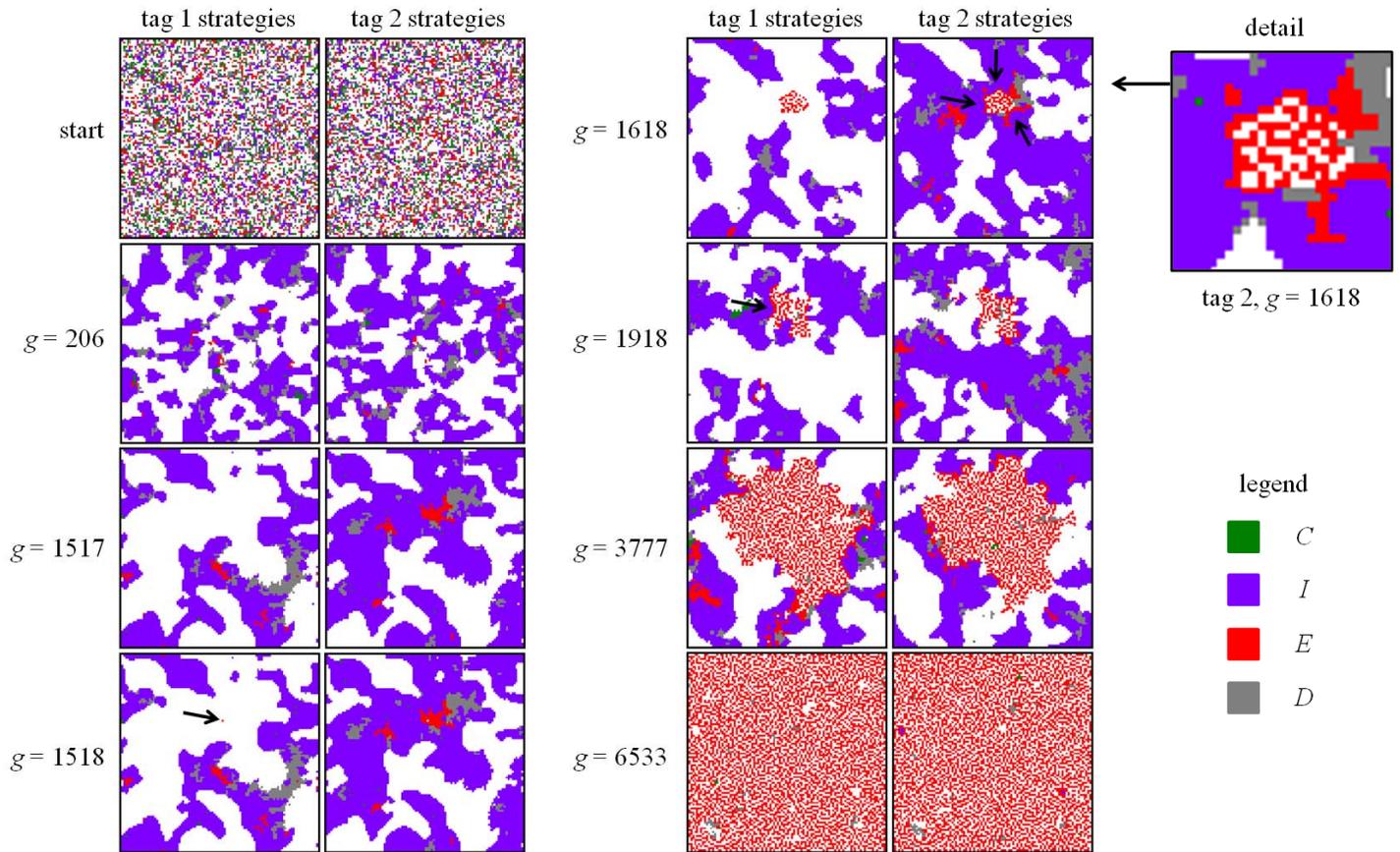
movie name	file name (hyperlink)	mutation rate ( $\mu$ )	cost-benefit ratio of mutual cooperation ( $r$ )	starting conditions (mixed or ancestral)
Movie D.1 <sup>a</sup>	<a href="#">SM01.mov</a>	$10^{-4}$	0.12	mixed
Movie D.2 <sup>b</sup>	<a href="#">SM02.mov</a>	$10^{-4}$	0.19	mixed
Movie D.3 <sup>c</sup>	<a href="#">SM03.mov</a>	$10^{-4}$	0.45	mixed
Movie D.4	<a href="#">SM04.mov</a>	$10^{-4}$	0.80	mixed
Movie D.5	<a href="#">SM05.mov</a>	0	0.15	mixed
Movie D.6	<a href="#">SM06.mov</a>	0	0.45	mixed
Movie D.7	<a href="#">SM07.mov</a>	$10^{-2}$	0.15	mixed
Movie D.8	<a href="#">SM08.mov</a>	$10^{-2}$	0.45	mixed
Movie D.9	<a href="#">SM09.mov</a>	$10^{-4}$	0.12	ancestral
Movie D.10	<a href="#">SM10.mov</a>	$10^{-4}$	0.19	ancestral
Movie D.11	<a href="#">SM11.mov</a>	$10^{-4}$	0.45	ancestral
Movie D.12	<a href="#">SM12.mov</a>	$10^{-4}$	0.80	ancestral

<sup>a</sup> same parameter values as shown in Fig. 4 in the main text

<sup>b</sup> same parameter values as shown in Fig. 3 in the main text

<sup>c</sup> same parameter values as shown in Fig. 2 in the main text

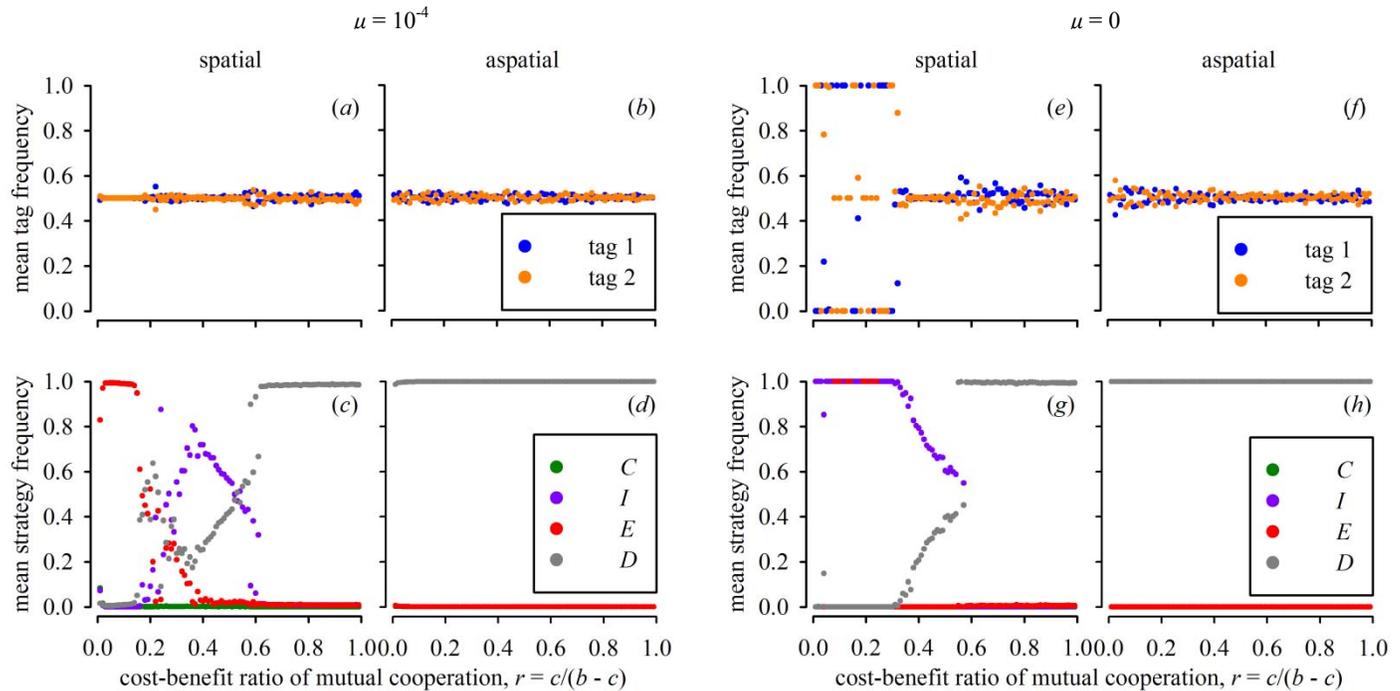
Figure D.1 shows a typical invasion sequence of a mixed-tag aggregation of  $E$ -strategists in a two-tag system when the cost-benefit ratio of mutual cooperation ( $r$ ) is low.



**Fig. D.1.** Typical invasion sequence of a coalition of extra-tag cooperators in a two-tag system. Panels show the spatial arrangement of tags (columns) and strategies ( $C$  = green,  $I$  = purple,  $E$  = red,  $D$  = grey) at key intervals (rows;  $g$  = number of model generations) for parameter values  $r = 0.12$  and  $\mu = 10^{-4}$ . Each pair of panels within a row is complementary, and represents the strategies adopted by Tag 1 (left) and Tag 2 (right); white regions in the lattices of one column indicate areas occupied by the other column's tag. In this example, at the start the tags and strategies are well mixed, but the results are similar when starting from all unconditional defectors of a single tag (see Appendix F and Movie D.9). After several generations ( $g = 206$ ),  $I$ - and  $D$ -strategists of both tags spontaneously self-organize into local aggregations that occupy most of the lattice; this exemplifies the traditional green-beard phenomenon.  $I$ -strategists of one tag can invade  $D$ -strategists of the other tag, but they are simultaneously invaded by unconditional defectors of their own tag, resulting in a non-transitive invasion loop whereby  $I_1 \rightarrow D_1 \rightarrow I_2 \rightarrow D_2 \rightarrow I_1$ . At the same time, within a single-tag aggregation of  $I$ -strategists, same-tagged  $E$ -strategists are situational defectors, and therefore can also invade (red clumps in  $g = 1517$ ). Eventually, an  $E_1$  mutant arises in a clump of  $E_2$  individuals ( $g = 1518$ , arrow), or vice versa. This results in a mutually miscible, mixed-tag aggregation of extra-tag cooperators that can invade the lattice, provided the cost-benefit ratio of mutual cooperation ( $r$ ) is not too high (see Fig. 1 in the main text). The mixed-tag aggregation is able to invade because regardless of which tag borders it, it always has a local supply of same-tagged traitors available to exploit more loyal strategies ( $g = 1618$  and  $1918$ , arrows). At low values of  $r$ , the traitorous  $E$ -strategists continue to spread ( $g = 3777$ ) until they eventually come to dominate the entire population ( $g = 6533$ ). At slightly greater values of  $r$ , extra-tag cooperators coexist in evolutionary time with intra-tag cooperators and unconditional defectors with whom they engage in non-transitive spatial invasibility. See Movies D.1 and D.2 for a dynamical view of these two modes of invasion by traitorous extra-tag cooperators.

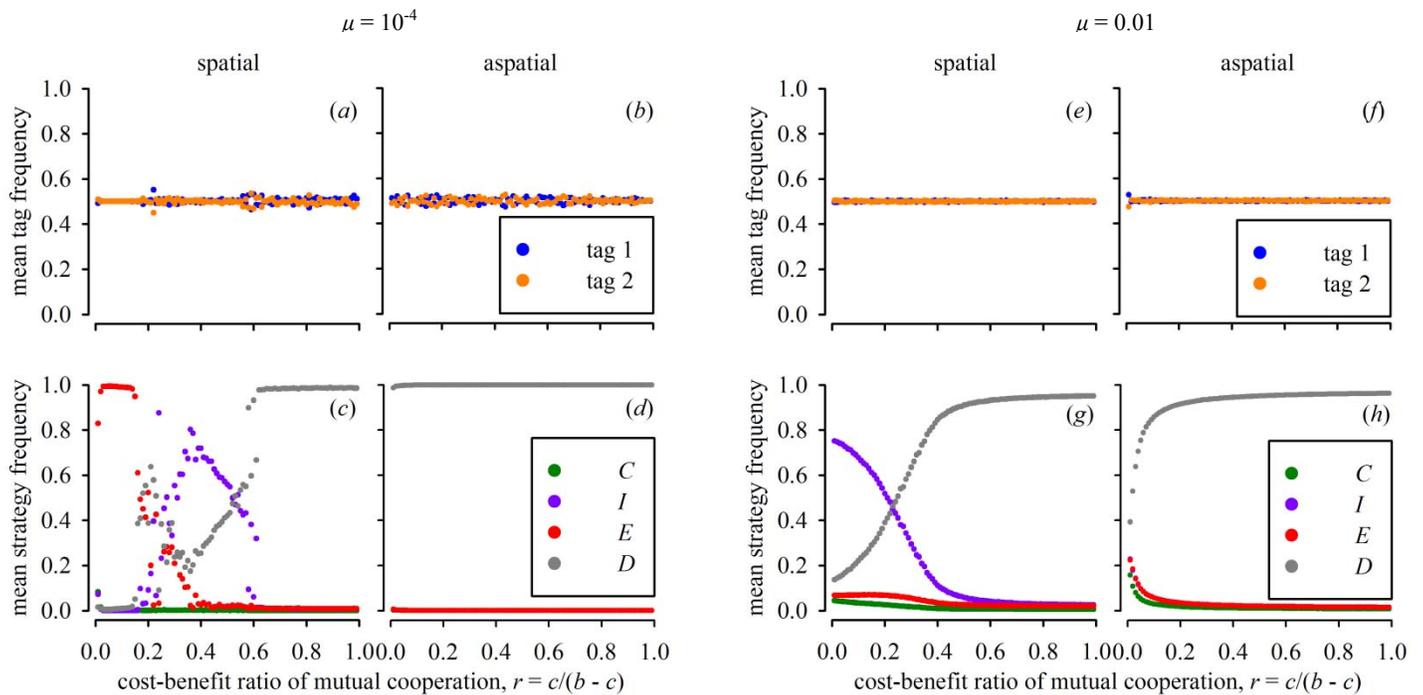
## Appendix E: Mutation and the evolution of traitorousness

Mutation is critical for the consistent evolution of traitorousness because it causes extra-tag cooperators ( $E$ ) bearing different tags to occur in the same neighbourhood, a precondition for their ability to spread locally. Only a small degree of mutation is necessary; the main text demonstrates how a single mutant per generation (on average) can lead to the quasi-stable or cyclic dominance of  $E$ -strategists in spatially structured populations. In the absence of mutation,  $E$  only rarely comes to dominate spatial populations (Fig. E.1).



**Fig. E.1.** Mutation and the evolution of traitorousness: Low versus no mutation. Panels show the average tag or strategy frequencies (the latter cumulative across both tags) between 15- and 20-thousand model generations for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and mutation rates of  $10^{-4}$  (a-d) or 0 (e-h); note that (a-d) recapitulates Fig. 1 from the main text. At the start, the tags and strategies were present in equal abundance and were well mixed. Low mutation (one mutant per model generation, on average) has little effect in the aspatial version of the model (compare (b, d) with (f, h)); in all cases  $D$  (grey) dominates (d, h) leading to the neutral coexistence of Tag 1 (blue) and Tag 2 (orange) (b, f). However, low mutation has a strong effect in the spatial version of the model (compare (a, c) with (e, g)). In particular, when there is no mutation there is no value of  $r$  that allows  $E$  (red) to consistently dominate or to engage in cyclic invasion dynamics with  $D$  and  $I$  (purple); rather, at low  $r$ ,  $I$  typically dominates, while  $E$  only rarely dominates (g). This rare dominance is attributable to spatial stochasticity (based on the well-mixed initial conditions) that occasionally allows mixed-tag groups of  $E$ -strategists to form, even in the absence of mutation. Interestingly, mutation also preserves tag coexistence at low  $r$  (compare (a), (e)): when the lack of mutation allows  $I$  to dominate (g), this is accompanied by the extinction of one of the two tags (e). See Movies D.5, D.6.

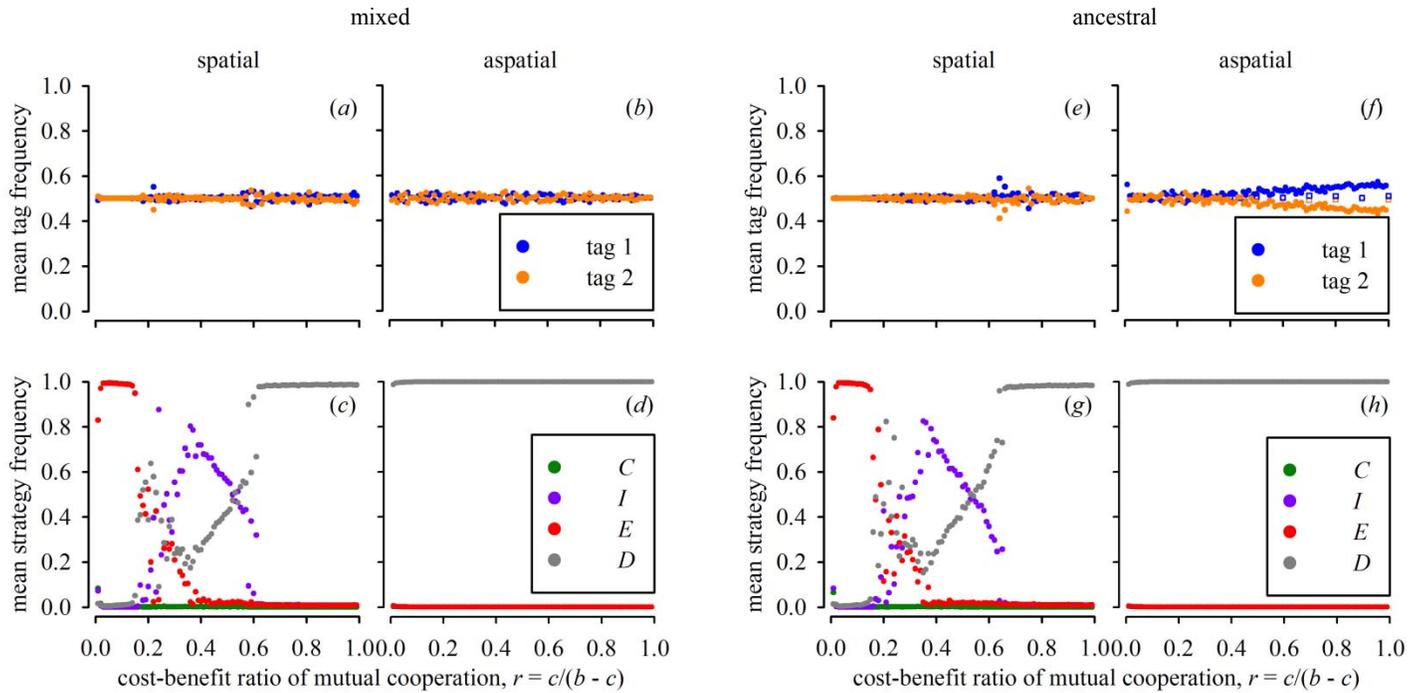
When the mutation rate is very high (e.g., one-hundred times greater than that discussed in the main text), traitorousness does not evolve to dominate populations (Fig. E.2). In this case, the high mutation causes new  $D$  mutants to undermine mixed-tag groups of  $E$ -strategists before they can grow.



**Fig. E.2.** Mutation and the evolution of traitorousness: Low versus high mutation. Panels show the average tag or strategy frequencies (the latter cumulative across both tags) between 15- and 20-thousand model generations for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and mutation rates of  $10^{-4}$  (a-d) or 0.01 (e-h); note that a-d recapitulates Fig. 1 from the main text. At the start, the tags and strategies were present in equal abundance and were well mixed. In the aspatial version of the model, high mutation allows  $C$  (green),  $I$  (purple), and  $E$  (red) to coexist at very low frequencies with  $D$  (grey) when  $r$  is very low (h). This is due to a shift in the mutation-selection balance, allowing non- $D$  strategies to persist at low  $r$  when the mutation rate is high (compare (h), (d)). In both cases, Tag 1 (blue) and Tag 2 (orange) coexist (b, f). In the spatial version of the model, high mutation undermines the ability of mixed-tag groups of  $E$ -strategists to grow. Hence, when  $r$  is low,  $I$  and  $D$  comprise most of the population with  $C$  and  $E$  persisting at very low levels (g); at no value of  $r$  does  $E$  dominate, as in the case of low mutation (compare c and g). See Movies D.7 and D.8.

## Appendix F: Different starting conditions in two-tag models

The simulations described in the main text started with a random distribution of tags and strategies. The results are similar when starting from what might be described as an ‘ancestral’ condition, i.e., when all individuals are unconditional defectors ( $D$ ) bearing the same tag (Fig. F.1).

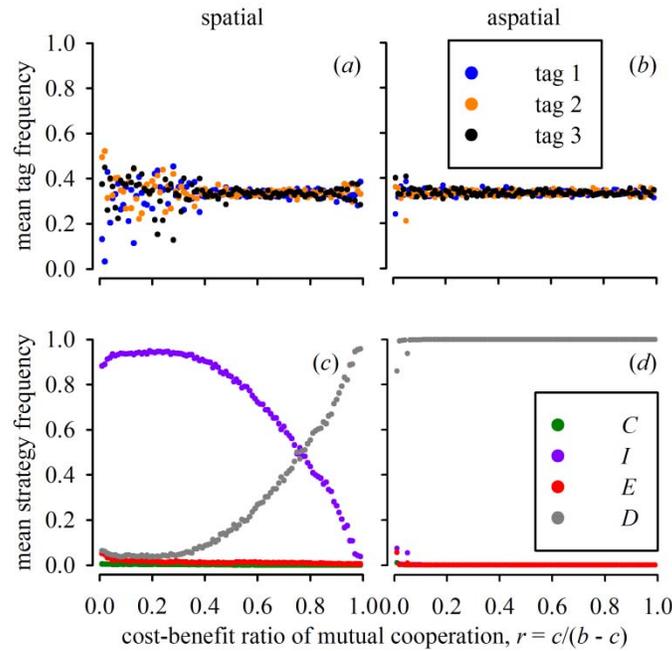


**Fig. F.1.** Different starting conditions. Panels show the average tag or strategy frequencies (the latter cumulative across both tags) between 15- and 20-thousand generations (closed circles) for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and a mutation rate of  $10^{-4}$ . In (a-d) (‘mixed’), the tags and strategies were initially present in equal abundance and were well mixed; note that a-d recapitulates Fig. 1 from the main text. In (e-h) (‘ancestral’), the model was started with only  $D_1$  individuals – i.e., unconditional defectors (grey) of Tag 1 (blue). The same evolutionary endpoint was reached for both starting conditions (compare (a-d) with (e-h)), although this endpoint took longer to reach in the aspatial ancestral version of the model, due to the slow action of mutation in equalizing tag frequencies (f; open squares show the average tag frequencies between 35- and 40-thousand generations for increments of  $r$  of 0.1). This delay increased with increasing values of  $r$  (f). This is because at low  $r$ , non- $D$  strategies are weeded out slowly, and thus there are multiple potential ‘indirect mutational paths’ between  $D_1$  and  $D_2$ . On the other hand, at high  $r$ , non- $D$  strategies are eliminated almost immediately, meaning that only  $D_1$  and  $D_2$  mutants can persist, thereby slowing the dynamics. See Movies D.9-D.12.

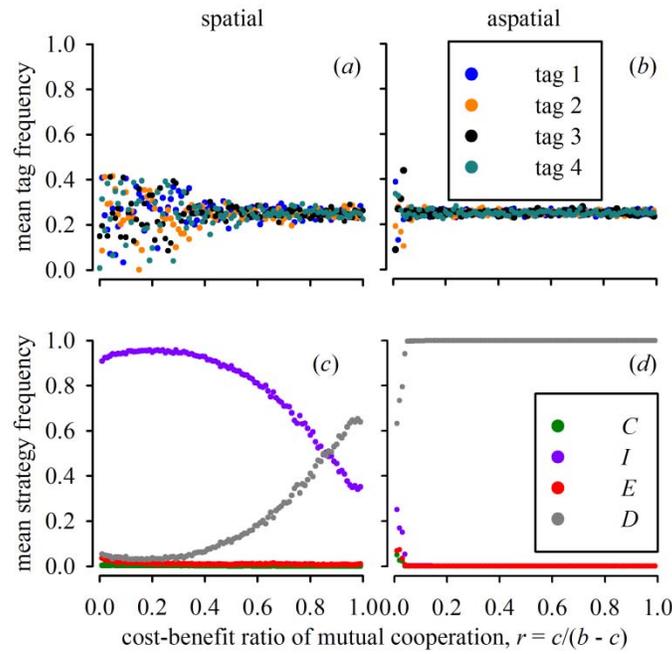
### Appendix G: Three-, four-, and five-tag models

In contrast to spatial two-tag models (see Fig. 1, main text), extra-tag cooperation does not evolve in spatial three-, four- and five-tag models (Figs. G.1, G.2, G.3). Rather, *I*-strategists dominate at relatively low cost-benefit ratios of mutual cooperation, *I*- and *D*-strategists coexist at intermediate cost-benefit ratios, and *D*-strategists dominate at relatively high cost-benefit ratios. Therefore, the presence of multiple tags appears to make green-beard cooperation relatively immune to traitorousness. Moreover, the more tags that are present in a population, the greater the range of cost-benefit ratios for which *I*-strategists dominate (Figs. G.1c, G.2c, G.3c).

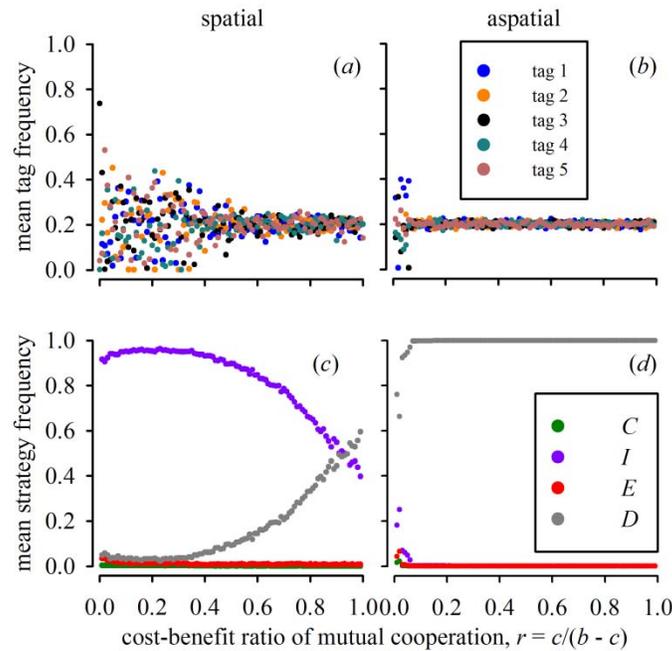
These results are not due solely to mutant limitation (see discussion in §3.3, main text). Figure G.4 shows that even when the lattice is started with only  $E_1$ ,  $E_2$ , and  $E_3$  individuals (and rare mutation), *I*-strategists still take over at relatively low cost-benefit ratios, and traitorousness never thrives. This is despite the fact that three-tag  $E_iE_jE_k$  aggregations (for example) perform similarly well as  $E_iE_j$  aggregations against  $C_i$ ,  $I_i$ , and  $D_i$  when there is no mutation (compare Figs. B.1 and G.5).



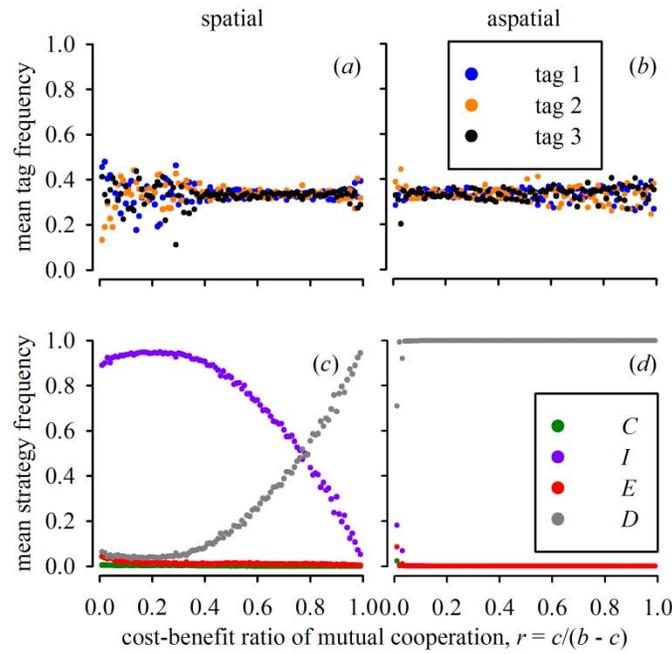
**Fig. G.1.** Three-tag models. Panels show the average tag or strategy frequencies (the latter cumulative across all tags) between 35- and 40-thousand model generations for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and a mutation rate of  $10^{-4}$ . At the start, the tags and strategies were present in equal abundance and were well mixed. Spatial and aspatial versions of the model are shown.



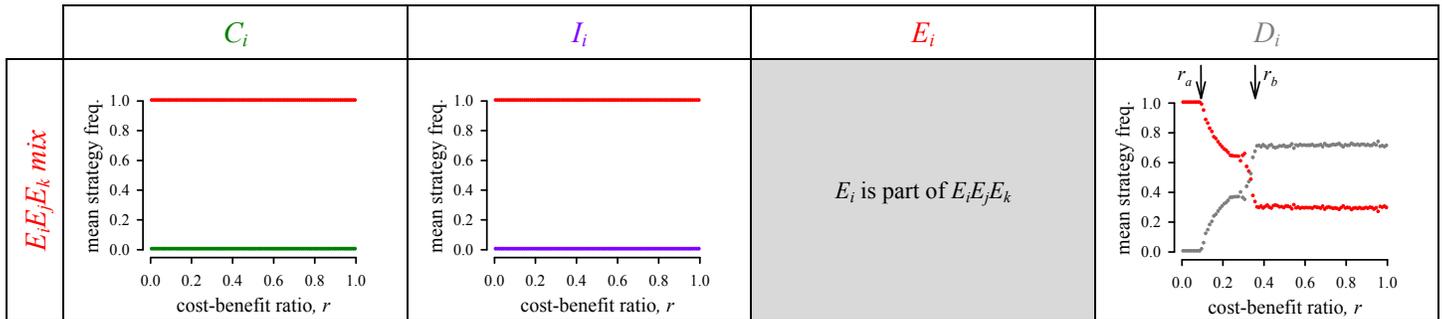
**Fig. G.2.** Four-tag models. Panels show the average tag or strategy frequencies (the latter cumulative across all tags) between 35- and 40-thousand model generations for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and a mutation rate of  $10^{-4}$ . At the start, the tags and strategies were present in equal abundance and were well mixed. Spatial and aspatial versions of the model are shown.



**Fig. G.3.** Five-tag models. Panels show the average tag or strategy frequencies (the latter cumulative across all tags) between 35- and 40-thousand model generations for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and a mutation rate of  $10^{-4}$ . At the start, the tags and strategies were present in equal abundance and were well mixed. Spatial and aspatial versions of the model are shown.



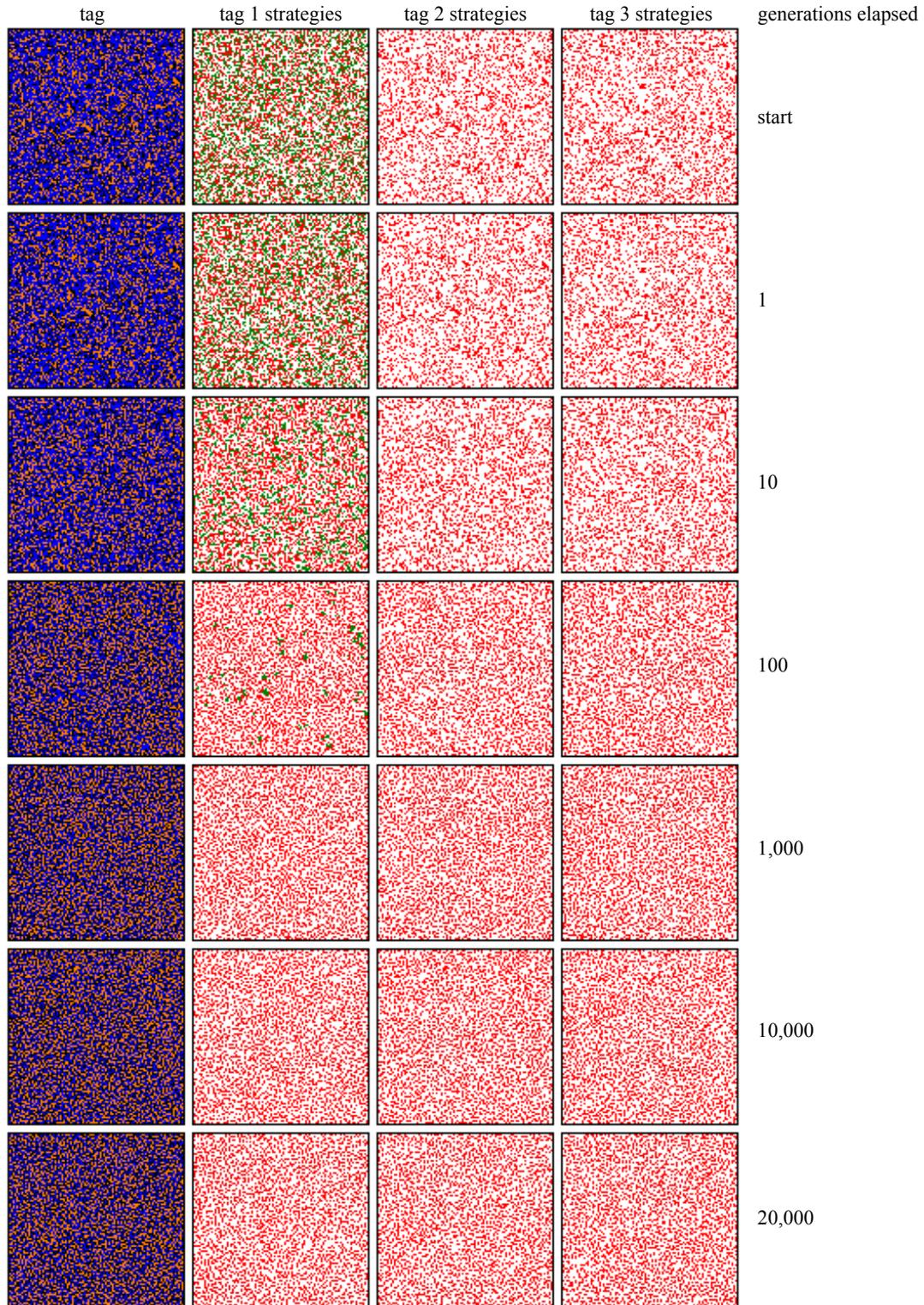
**Fig. G.4.** Three-tag models, starting with only  $E$ -strategists. Panels show the average tag or strategy frequencies (the latter cumulative across all tags) between 15- and 20-thousand model generations for cost-benefit ratios  $0 < r < 1$  in increments of 0.01 and a mutation rate of  $10^{-4}$ . At the start,  $E$ -strategists belonging to the tree tags were present in equal abundance and were well mixed. Spatial and aspatial versions of the model are shown. Note that the results are very similar to those presented in Fig. G.1.



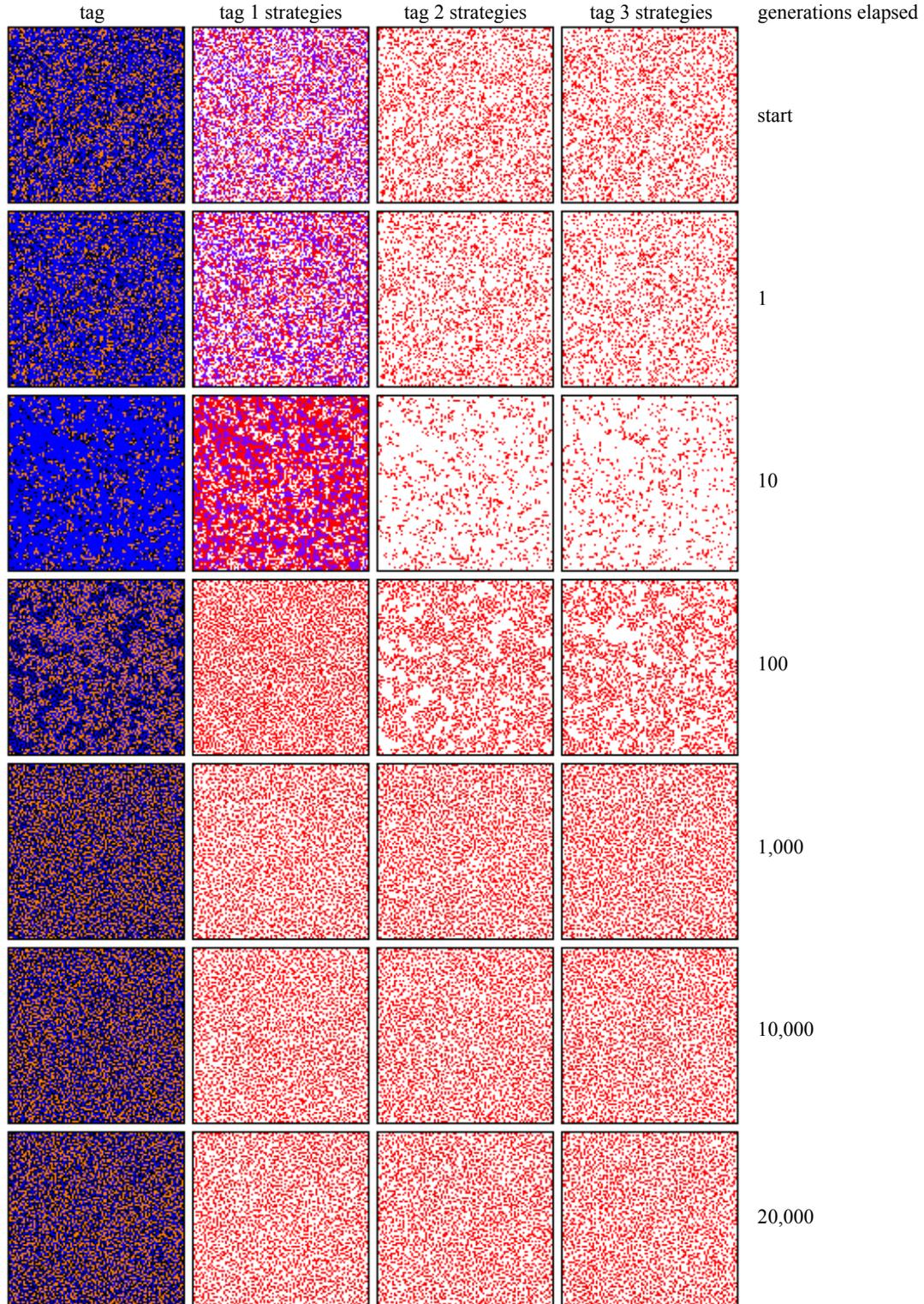
**Fig. G.5.** Spatial invasion outcomes when there are four strategy-by-tag types are present (i.e., a mixed-tag aggregation,  $E_i E_j E_k$ , and one other type, given by the column header) and there is no mutation. Each panel shows the average strategy frequency between 15- and 20-thousand generations of the spatial model, for populations of 10,000 individuals, and cost benefit ratios of  $0 < r < 1$  in increments of 0.01. Initially there was a random arrangement of strategies. Line colours correspond to the colours of the row and column headers, with  $E_i$ ,  $E_j$ , and  $E_k$  shown in aggregate.  $E_i$  is part of  $E_i E_j E_k$  and is therefore subsumed by it.

Figures G.6-G.8 show snapshots of example simulations of the main outcomes of spatial interactions between tag-by-strategy types, in the absence of mutation, as described for three-tag systems in Fig. G.5:

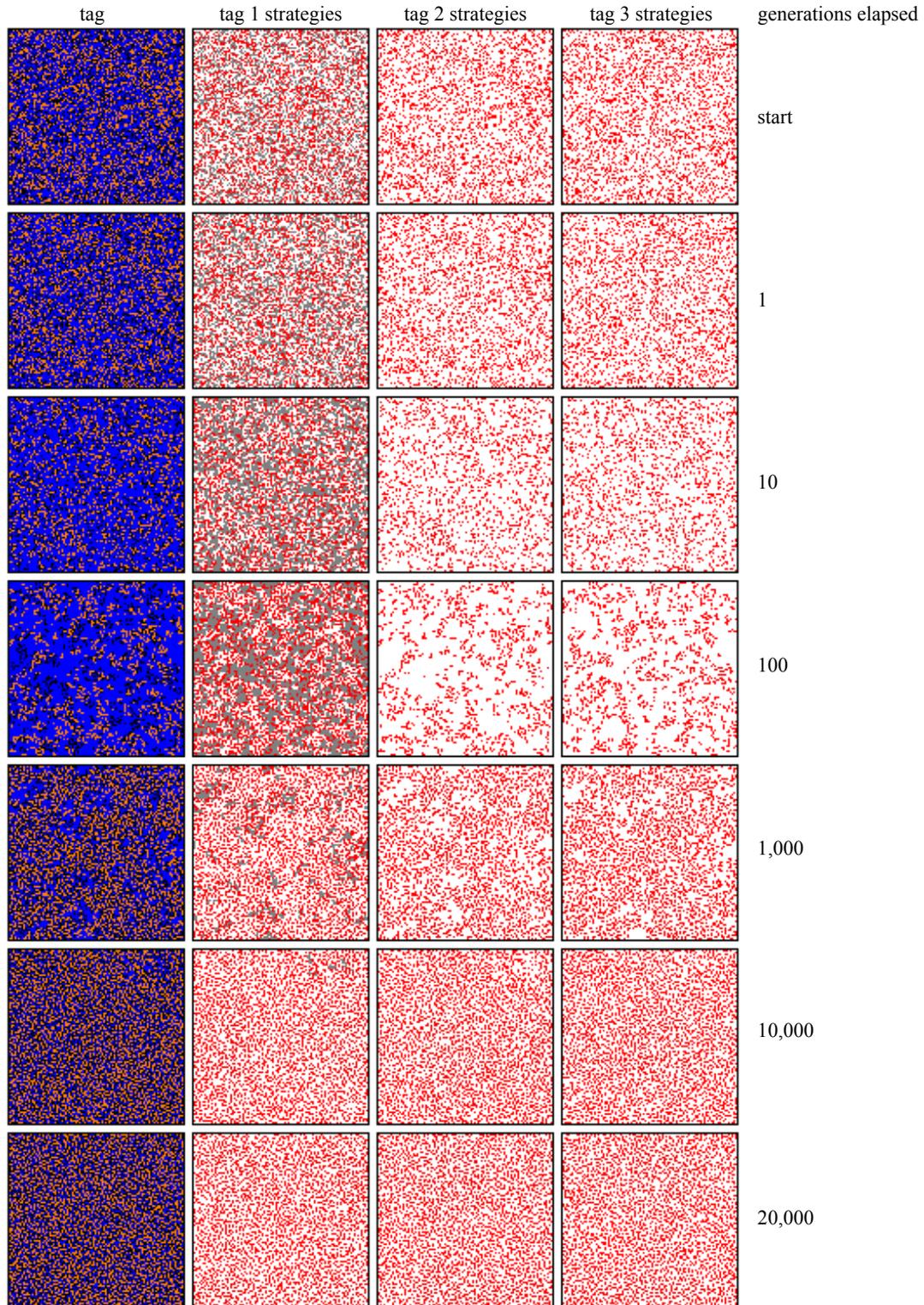
- Figure G.6 gives an example of a mixed-tag aggregation of  $E_iE_jE_k$  overtaking  $C_i$ .
- Figure G.7 gives an example of a mixed-tag aggregation of  $E_iE_jE_k$  overtaking  $I_i$ .
- Figure G.8 gives an example of a mixed-tag aggregation of  $E_iE_jE_k$  overtaking  $D_i$  at low  $r$  (i.e.,  $r < r_a \approx 0.10$ ).
- Figure G.9 gives an example of a mixed-tag aggregation of  $E_iE_jE_k$  coexisting with  $D_i$  at intermediate  $r$  (i.e.,  $r_a < r < r_b \approx 0.37$ ).
- Figure G.10 gives an example of  $D_i$  purging  $E_j$  and  $E_k$  from a mixed-tag aggregation of  $E_iE_jE_k$  at high  $r$  (i.e.,  $r > r_b$ ).



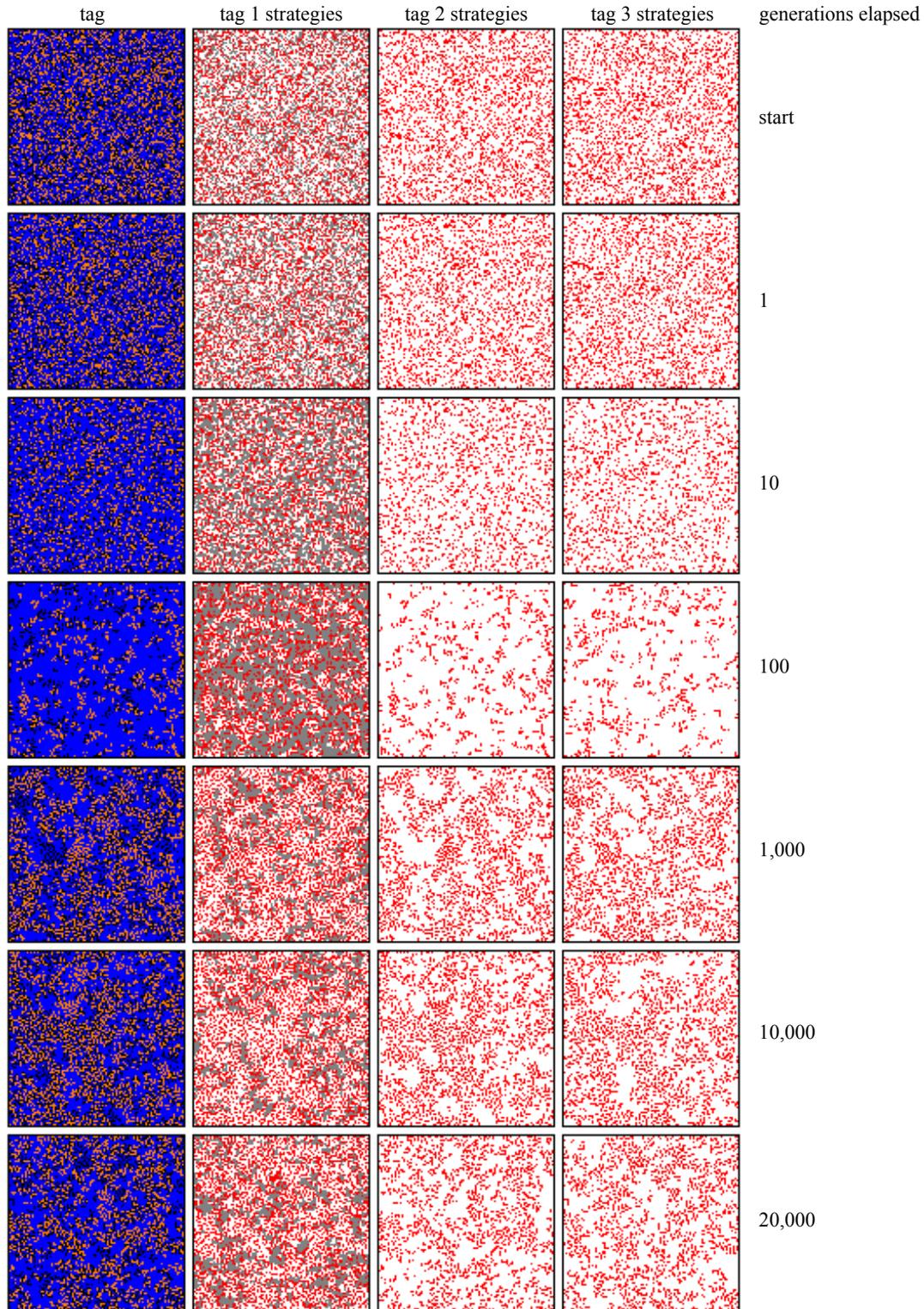
**Fig. G.6.** Snapshots of spatial lattice model for  $E_i E_j E_k$  versus  $C_i$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as four corresponding panels for seven generations (start-20,000). The left panel shows the spatial positions of Tag  $i = 1$  (blue), Tag  $j = 2$  (orange), and Tag  $k = 3$  (black). The three right-most panels show the spatial positions of the strategies of Tag 1 ( $E = \text{red}$ ,  $C = \text{green}$ ), Tag 2 ( $E = \text{red}$ ), and Tag 3 ( $E = \text{red}$ ) individuals.  $E_i E_j E_k$  overtakes  $C_i$  irrespective of  $r$ .



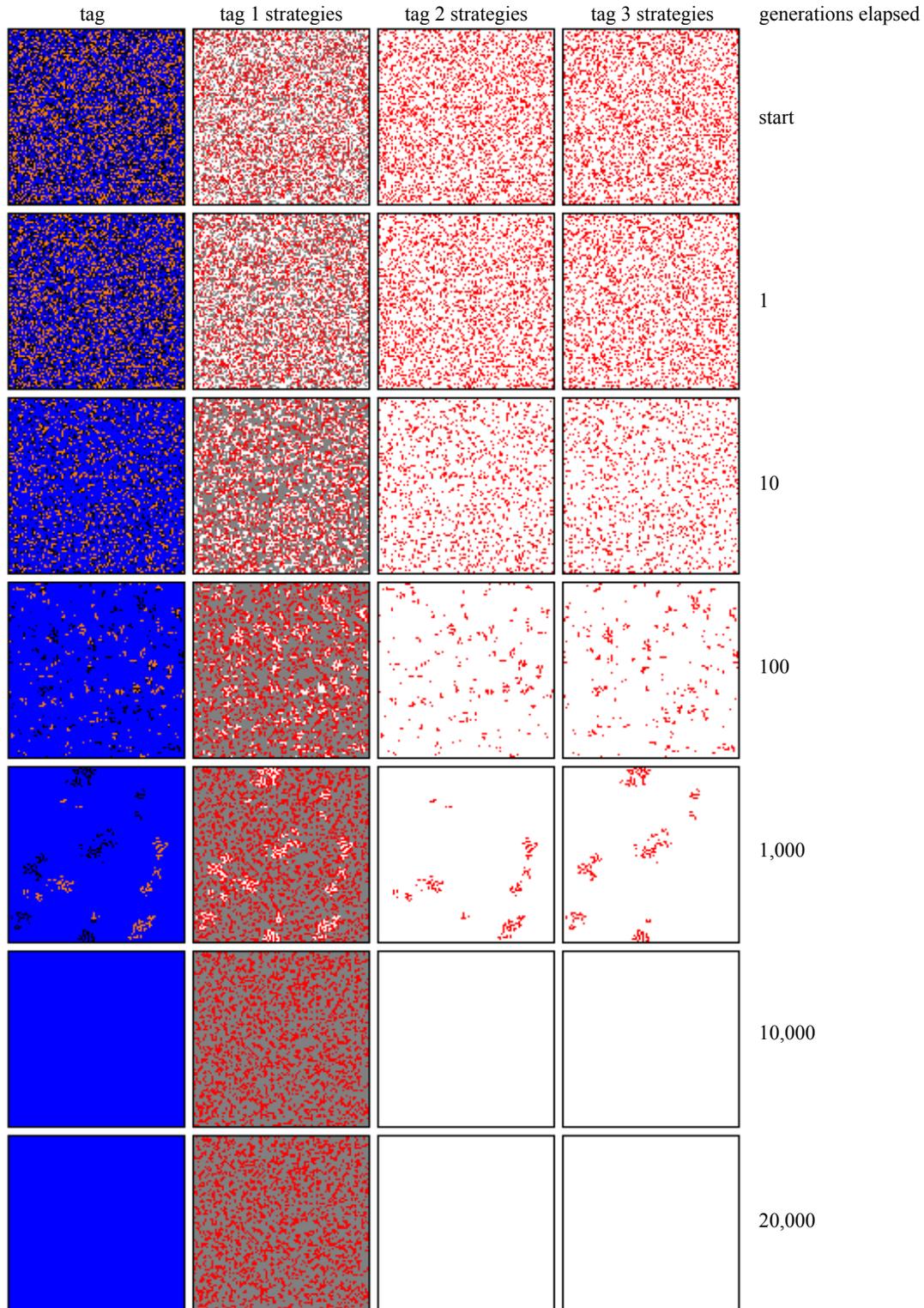
**Fig. G.7.** Snapshots of spatial lattice model for  $E_i E_j E_k$  versus  $I_i$  ( $\mu = 0$ ,  $r = 0.50$ ). The lattice is shown as four corresponding panels for seven generations (start-20,000). The left panel shows the spatial positions of Tag  $i = 1$  (blue), Tag  $j = 2$  (orange), and Tag  $k = 3$  (black). The three right-most panels show the spatial positions of the strategies of Tag 1 ( $E = \text{red}$ ,  $I = \text{purple}$ ), Tag 2 ( $E = \text{red}$ ), and Tag 3 ( $E = \text{red}$ ) individuals.  $E_i E_j E_k$  overtakes  $I_i$  for all values of  $r$ .



**Fig. G.8.** Snapshots of spatial lattice model for  $E_i E_j E_k$  versus  $D_i$  ( $\mu = 0$ ,  $r = 0.09$ ). The lattice is shown as four corresponding panels for seven generations (start-20,000). The left panel shows the spatial positions of Tag  $i = 1$  (blue), Tag  $j = 2$  (orange), and Tag  $k = 3$  (black). The three right-most panels show the spatial positions of the strategies of Tag 1 ( $E = \text{red}$ ,  $D = \text{grey}$ ), Tag 2 ( $E = \text{red}$ ), and Tag 3 ( $E = \text{red}$ ) individuals.  $E_i E_j E_k$  overtakes  $D_i$  for low values of  $r$ .



**Fig. G.9.** Snapshots of spatial lattice model for  $E_i E_j E_k$  versus  $D_i$  ( $\mu = 0$ ,  $r = 0.15$ ). The lattice is shown as four corresponding panels for seven generations (start-20,000). The left panel shows the spatial positions of Tag  $i = 1$  (blue), Tag  $j = 2$  (orange), and Tag  $k = 3$  (black). The three right-most panels show the spatial positions of the strategies of Tag 1 ( $E = \text{red}$ ,  $D = \text{grey}$ ), Tag 2 ( $E = \text{red}$ ), and Tag 3 ( $E = \text{red}$ ) individuals.  $E_i E_j E_k$  and  $D_i$  coexist for intermediate values of  $r$ .



**Fig. G.10.** Snapshots of spatial lattice model for  $E_i E_j E_k$  versus  $D_i$  ( $\mu = 0$ ,  $r = 0.38$ ). The lattice is shown as four corresponding panels for seven generations (start-20,000). The left panel shows the spatial positions of Tag  $i = 1$  (blue), Tag  $j = 2$  (orange), and Tag  $k = 3$  (black). The three right-most panels show the spatial positions of the strategies of Tag 1 ( $E = \text{red}$ ,  $D = \text{grey}$ ), Tag 2 ( $E = \text{red}$ ), and Tag 3 ( $E = \text{red}$ ) individuals. At high values of  $r$ ,  $D_i$  purges  $E_j$  and  $E_k$  from the mixed-tag aggregation  $E_i E_j E_k$ , leading to a standoff between  $D_i$  and  $E_i$ , caused by the fact that all the interactions between  $D_i$  and  $E_i$  result in mutual defection.