

TI 2019-073/I
Tinbergen Institute Discussion Paper

The cancellation effect at the group level

Aslihan Akdeniz^{1,2}

Matthijs van Veelen²

¹ Department of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands

² Tinbergen Institute, The Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

The cancellation effect at the group level

Ashhan Akdeniz^{1,2,*} and Matthijs van Veelen^{1,2}

¹*Department of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands.*

²*Tinbergen Institute, The Netherlands.*

*e-mail: a.akdeniz@uva.nl

Abstract

Group selection models combine selection pressure at the individual level with selection pressure at the group level (Sober and Wilson, 1998; Traulsen and Nowak, 2006; Wilson and Wilson, 2007; Boyd and Richerson, 2009; Simon, 2010; Simon et al., 2013; Luo, 2014; van Veelen et al., 2014; Luo and Mattingly, 2017). Cooperation can be costly for individuals, but beneficial for the group, and therefore, if individuals are sufficiently much assorted, and cooperators find themselves in groups with disproportionately many other cooperators, cooperation can evolve. The existing literature on group selection generally assumes that competition between groups takes place in a well-mixed population of groups, where any group competes with any other group equally intensely. Competition between groups however might very well occur locally; groups may compete more intensely with nearby than with far-away groups. We show that if competition between groups is indeed local, then the evolution of cooperation can be hindered significantly by the fact that groups with many cooperators will mostly compete against neighbouring groups that are also highly cooperative, and therefore harder to outcompete. The existing empirical method for determining how conducive a group structured population is to the evolution of cooperation also implicitly assumes global between group competition, and therefore gives (possibly very) biased estimates.

There is a wide variety of positions on the role of group selection in human evolution. One end of the spectrum considers group selection to be a key ingredient of human evolution (Sober and Wilson, 1998; Wilson and Wilson, 2007; Haidt, 2012). The other side suggests that “group selection has no useful role to play in psychology or social science” (Pinker, 2015). In this paper we will not resolve this controversy, nor take a position in this debate, but what we will do is consider a crucial element that has been missing, both from the current group selection models, and from the current empirical approach to establishing how conducive group structure is to the evolution of cooperation.

The defining characteristic of a group selection model is that it captures the opposing effects of selection at the individual level, where defectors do better than cooperators within groups, and selection at the group level, where groups with more cooperators do better than groups with fewer cooperators. The existing models within the group selection literature all share the property that competition between groups happens globally; all groups compete with all other groups equally intensely (Traulsen and Nowak, 2006; Boyd and Richerson, 2009; Simon, 2010; Simon et al., 2013; Luo, 2014; Luo and Mattingly, 2017; van Veelen et al., 2014). This is a useful simplification if the aim is to illustrate the possibility of a tug of war between the different levels of selection. It may however not always be particularly realistic. Groups themselves typically live in a structured population of groups, where neighbouring groups compete with each other more than they do with groups that are farther away. Local dispersal would then imply that groups with many cooperators are typically

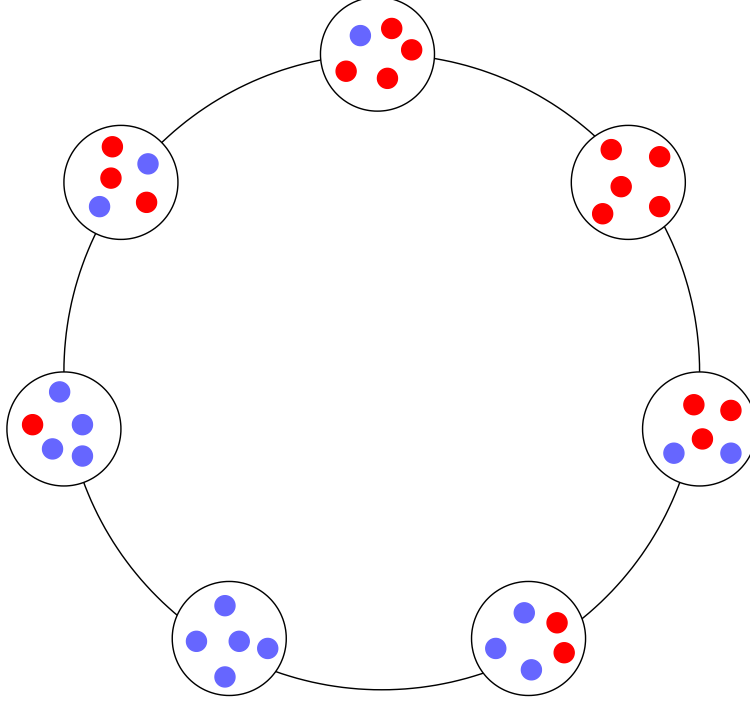


Figure 1: An example of a population state on a cycle with $m = 7$ groups of $n = 5$ individuals each. The blue dots indicate cooperators and the red dots indicate defectors.

surrounded by groups that also contain many cooperators, compared to the groups that surround groups with many defectors. More cooperative groups therefore might also be subject to more intense competition at the group level. This can significantly dampen the benefits of being a cooperative group, which, in turn, affects the balance between selection at the individual and at the group level. In models without group structure a similar phenomenon, but then at the individual level, is called the cancellation effect (Wilson et al., 1992; Taylor, 1992a;b). We show that the cancellation effect also exists at the group level, where it plays out in a more complex way, and that it can make a sizable difference for the conditions under which cooperation can evolve by group selection. This also has empirical implications. The current standard approach to determining how large the benefit to the group should be, compared to the cost to the individual, for cooperation to evolve by group selection implicitly assumes global between group competition (Weir and Cockerham, 1984; Crow and Aoki, 1984; Aoki and Nozawa, 1984; Bowles, 2006; Bell et al., 2009; Langergraber et al., 2011; Walker, 2014). If competition between groups is not global, but at least to some extent local, then this procedure paints too positive a picture of how favourable conditions are for the evolution of cooperation by group selection.

In order to study the difference between global and local group competition, we consider a stylized model, in which m groups consisting of n individuals live on a cycle (Fig. 1). Individuals can either be a cooperator (C) or a defector (D). In every time period, one of three types of events will happen: individual reproduction, group reproduction, or migration. These events happen with probabilities p , q and r , respectively, where $p + q + r = 1$. We compare two different processes for group reproduction, one with local and one with global between group competition.

If an individual reproduction event occurs, first a random group is selected, where all groups have equal

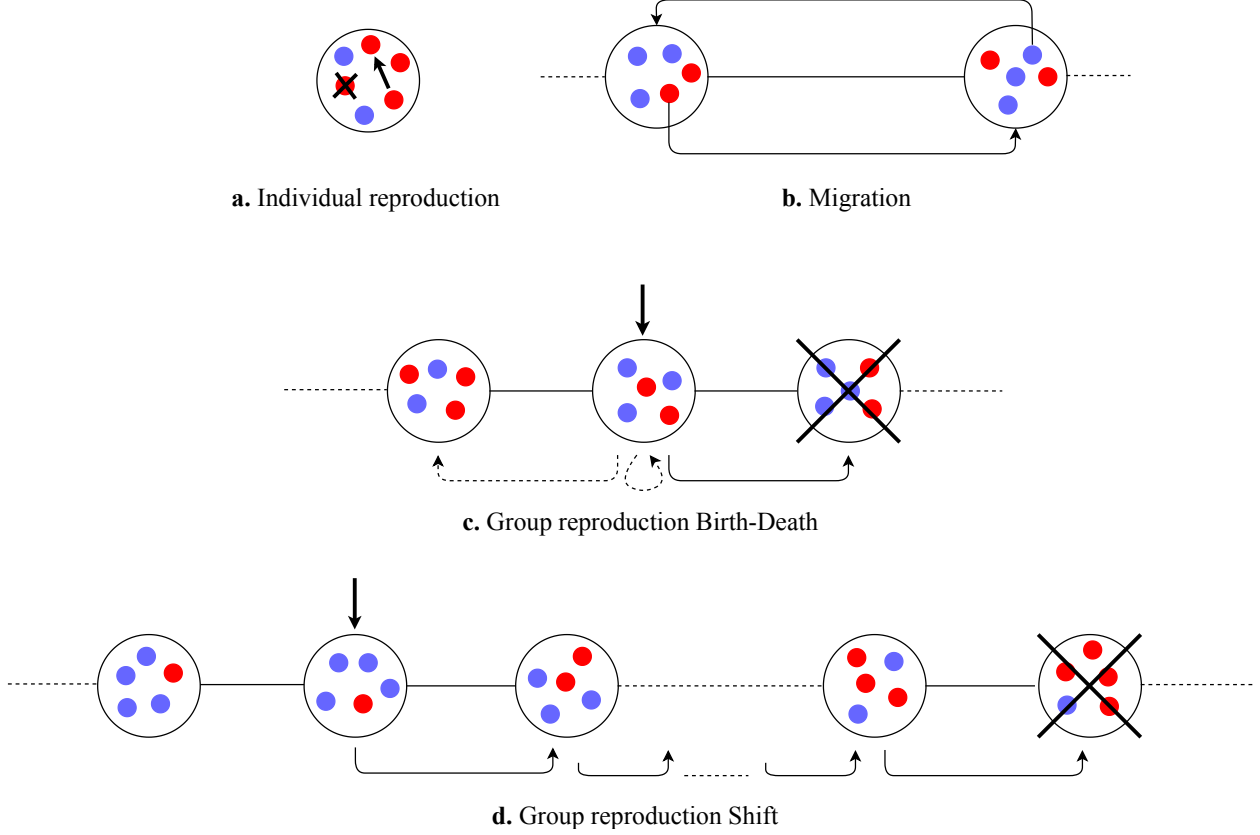


Figure 2: (a) At an individual reproduction event, one individual reproduces, and one individual within the same group dies. In this example, one defector is chosen for reproduction, and another defector is chosen to die, so the overall group composition has not changed. Defectors have a higher chance of being chosen for individual reproduction than cooperators do. (b) At a migration event, two individuals from neighbouring groups trade places. (c) In the Birth-Death (BD) process, the group that is chosen to reproduce produces an identical offspring group. This offspring group then replaces one of the neighbouring groups, or, with a small probability, it replaces the parent group itself. (d) In the Shift process, the group that is chosen to reproduce also produces an identical offspring group, but here any group can be chosen to die, including the parent group. If the parent group and the dying group are more than 1 position apart, all groups between them move over one position. In both processes, groups with many cooperators have a higher chance of being chosen for group reproduction than groups with many defectors.

probability of being chosen. Then an individual from the selected group is chosen to reproduce. Within the group, defectors get a payoff of 1 and cooperators get a payoff of $1 - c$. The intensity of selection at the individual level w_{ind} is then used to transform these payoffs to values f_C and f_D :

$$f_C = 1 - w_{ind}c \quad \text{and} \quad f_D = 1$$

The probabilities with which individuals are chosen for reproduction within the group are proportional to these values. Whenever an individual reproduces, an individual from the same group is chosen to die, where all individuals, including the parent, but excluding the offspring, are chosen with probability $\frac{1}{n}$ (Fig. 2a).

If a group reproduction event occurs, then one group is chosen to reproduce, and one group is chosen to die. The groups are numbered $i = 1, \dots, m$, and k_i is the number of cooperators in group i . These groups live on a cycle, so i and $i + 1$ are neighbouring groups, and so are groups 1 and m . The group payoff of group i is 1

plus b times the share of cooperators in the group. The intensity of selection at the group level w_{gr} is then used to transform these payoffs to values

$$g_i(k_i) = 1 + w_{gr} \frac{k_i}{n} b$$

We consider two update processes for group reproduction; Birth-Death (BD) and Shift. In both of them, first a group is chosen for reproduction, where each group's probability of being chosen is proportional to the values $g_i(k_i)$. With BD, the offspring group then replaces the left or the right neighbour of the parent group, both with probability $\frac{m-1}{2m}$, and it replaces its own parent group with probability $\frac{1}{m}$ (Fig. 2c). This makes competition at the group level local. With Shift, all groups, including the parent group, but excluding the offspring group, die with probability $\frac{1}{m}$. Unless the offspring group replaces the parent group, the new group is placed either to the right or to the left of the parent group, with equal probability, and every other group in between the parent group and the dying group moves over one spot (Fig. 2d). With Shift, every group is equally likely to die, irrespective of the composition of their neighbouring groups. Competition between groups is therefore global, as it is in the standard group selection models that have a well-mixed population of groups.

Finally, if a migration event happens, then a random pair of individuals from neighbouring groups trade places (Fig. 2b).

In the Supplementary Information, we derive critical b/c ratios, above which cooperators have a selective advantage. These conditions are derived in the limit of weak selection, where w_{ind} and w_{gr} are equal to each other, and both approach 0. With the BD process, cooperation is selected for if

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{m}{m-1} \frac{1-r_s}{1+(n-1)r_s-nr_1} \quad (1)$$

where r_s is the relatedness between two different individuals from the same group, and r_1 is the relatedness between two individuals from neighbouring groups. With the Shift process, cooperation is selected for if

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{1-r_s}{1+(n-1)r_s} \quad (2)$$

There are three differences between these two thresholds. The first is that Equation 1 has an $\frac{m}{m-1}$ term that is absent in Equation 2. For a large number of groups m , this term approaches 1, which means that for the cases we are most interested in, this does not contribute too much to the gap between the two b/c ratios. The second difference is that r_s is not the same across the two processes, even if everything else (that is: p , q , r , n and m) is equal. In the Supplementary Information, we also calculate how r_s depends on those five parameters for the two different processes, and it turns out that r_s is higher for BD than for Shift. Therefore, if it was not for the third difference, the critical b/c ratio would actually be *lower* for BD than for Shift. The third difference is that in the last fraction in the equation, there is a $-nr_1$ term in the denominator for BD, but not for Shift. With BD, groups replace their direct neighbours, and for the cancellation effect it is important how similar these direct neighbours are. That is why r_1 appears in the formula for BD, and not in the formula for Shift, where groups compete globally. The difference between the critical b/c ratio for BD and the one for Shift can range from negligibly small (both for migration rates close to 0 and for migration rates close to 1; Fig. 3c) to quite substantial (for not too small numbers of groups

m at intermediate migration rates r ; Fig. 3a and 3b). Simulations not in the limit of weak selection show a similar gap. The Supplementary Information also contains a mathematical proof that the threshold for BD is always higher than the one for Shift as soon as the number of groups exceeds 3. For $m = 2$ or $m = 3$, the two different update processes imply the same dynamic.

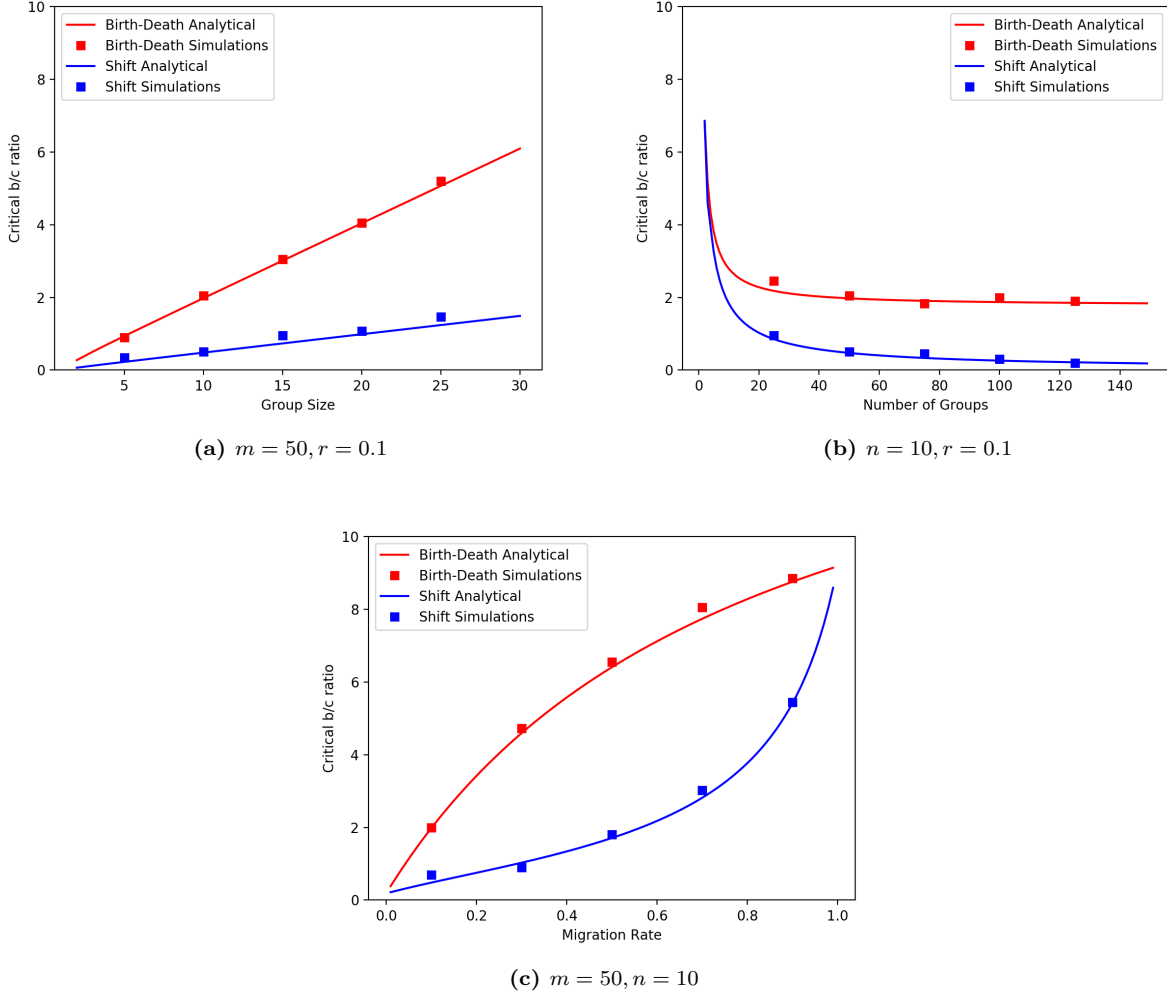


Figure 3: Critical b/c ratios in the limit of weak selection for BD (red lines) and Shift (blue lines), as well as simulation results at an intensity of selection of 0.1, both at the individual and at the group level ($w_{ind} = w_{gr} = 0.1$), for BD (red squares) and Shift (blue squares). In panels (a) and (b), one in every ten events is a migration event ($r = 0.1$). In panels (a) and (c), the number of groups is set to $m = 50$. In panels (b) and (c), the group size is set to $n = 10$. Probabilities p and q are chosen so that the average individual is as likely to die as a result of an individual reproduction event as it is to die from a group reproduction event under neutral selection: $p = (1 - r) \frac{1}{n+1}$ and $q = (1 - r) \frac{n}{n+1}$. Similar to a model without population structure at the group level (Traulsen and Nowak, 2006), larger group sizes (a) and larger migration rates (c) increase the critical b/c ratio, and larger numbers of groups (b) decrease it. The gap between BD and Shift is there for a range of group sizes, numbers of groups, and migration rates. The gap between the two processes disappears when the migration rate vanishes, in which case the dynamics are such that all groups are at within-group fixation almost all of the time (see the SI for why that makes the gap disappear). The gap also disappears when the migration rate is close to 1, and the whole population is shaken and stirred between any two reproduction events. Again the SI gives the complete argument why relatednesses r_s and r_1 being close to 0 not only means that the right hand sides of Equations (1) and (2) should be similar, but are actually the exact same in the limit of $r \rightarrow 1$.

These thresholds are derived using inclusive fitness. We can do this, because the effects that being a cooperator instead of a defector has on individual reproduction rates, and on individuals death rates, as well as the effects it has on reproduction and death rates of groups, satisfy generalized equal gains from switching in the limit of weak selection (van Veelen, 2018; van Veelen et al., 2017). It may be worth pointing out that the b and the c in those thresholds are model parameters, and that these translate to a variety of fitness effects. These fitness effects comprise all changes in birth and death rates of all individuals and groups as a result of an individual being a cooperator instead of a defector. Two important ones are the negative effect on the individual’s own birth rate, which is proportional to c , and the positive effect on the reproduction rate of the group it is in, which is proportional to b . These are the same for both update processes. All other effects on birth and death rates of individuals, as well as on the group reproduction rates, are also the same in both processes. The only thing that is different between the two processes is how being a cooperator affects the death rates of other groups. With BD, the neighbouring groups are the only ones that are replaced at a higher rate, while the death rates of all other groups remain unaffected. With Shift, every group, close by or far away, suffers an equally small increase in their death rate.

Our results show that in models of group selection, the evolution of cooperation can be quite a bit harder if between group competition is local instead of global. Our model captures two extremes; between group competition is completely local with the Birth Death process, and completely global with Shift, and the difference in critical b/c -ratios between them can be more than substantial. Our findings also clearly have empirical implications. The current, well-established approach in empirical studies concerning group selection is to measure F_{ST} ’s in order to determine how large the benefit to the group should be, compared to the cost to the individual, for cooperation to evolve (Weir and Cockerham, 1984; Crow and Aoki, 1984; Aoki and Nozawa, 1984; Bowles, 2006; Bell et al., 2009; Langergraber et al., 2011; Walker, 2014). The implicit assumption in that approach is that competition between groups is global (see the derivation in the SI). If competition between groups is not global, but at least to some extent local, then this procedure paints too positive a picture of how favourable conditions are for the evolution of cooperation by group selection. Moreover, if we want to determine how large the benefits to the group should be, relative to the costs to the individual, then measuring F_{ST} ’s is not enough; one would also need to determine how local or global between group competition is. While BD and Shift are two extremes, with competition between groups being entirely local versus completely global, realistic between-group competition may very well be somewhere in between. Finding out exactly how local or global group competition is, however, will not be easy. Section 9 in the SI discusses the statistical challenges, and also shows how in statistical models, if they incorrectly assume global group competition, the terms meant to measure fitness effects end up also absorbing the cancellation effect, and therefore give biased estimates of the size of the fitness effects.

References

- Benjamin Allen and Alex McAvoy. A mathematical formalism for natural selection with arbitrary spatial and genetic structure. *arXiv:1806.04717 [q-bio.PE]*, 2018.
- Benjamin Allen and Martin A. Nowak. Evolutionary shift dynamics on a cycle. *Journal of Theoretical Biology*, 311:28–39, 2012.
- Benjamin Allen and Corina E. Tarnita. Measures of success in a class of evolutionary models with fixed population size and structure. *Journal of Mathematical Biology*, 68:109–143, 2014.
- Benjamin Allen, Arne Traulsen, Corina E. Tarnita, and Martin A. Nowak. How mutation affects evolutionary games on graphs. *Journal of Theoretical Biology*, 299:97–105, 2012.
- Kenichi Aoki and Ken Nozawa. Average coefficient of relationship within troops of the japanese monkey and other primate species with reference to the possibility of group selection. *Primates*, 25(2):171–184, 1984.
- Adrian V. Bell, Peter J. Richerson, and Richard McElreath. Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences*, 106(42):17671–17674, 2009.
- Samuel Bowles. Group competition, reproductive leveling, and the evolution of human altruism. *Science*, 314(5805):1569–1572, 2006.
- Samuel Bowles. Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science*, 324(5932):1293–1298, 2009.
- Rob Boyd. Density-dependent mortality and the evolution of social interactions. *Animal Behaviour*, 30:972–982, 1982.
- Robert Boyd and Peter J. Richerson. Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3281–3288, 2009.
- James F. Crow and Kenichi Aoki. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences*, 81(19):6073–6077, 1984.
- Richard Durrett. *Probability models for DNA sequence evolution. Probability and its applications, 2nd Edition*. Springer, New York, 2008.
- Alan Grafen. Natural selection, kin selection and group selection. In J.R. Krebs and N.B. Davies, editors, *Behavioural Ecology, 2nd edn.*, pages 62–84. Blackwell, 1983.
- Alan Grafen. An inclusive fitness analysis of altruism on a cyclical network. *Journal of Evolutionary Biology*, 20(6):2278–2283, 2007.
- Jonathan Haidt. *The righteous mind: Why good people are divided by politics and religion*. Vintage, 2012.
- William D. Hamilton. The genetical evolution of social behaviour. I. *Journal of theoretical biology*, 7(1):1–16, 1964a.
- William D. Hamilton. The genetical evolution of social behaviour. II. *Journal of theoretical biology*, 7(1):17–52, 1964b.

- William D Hamilton. Selection of selfish and altruistic behaviour in some extreme models. In J.F. Eisenberg and W.S. Dillon, editors, *Man and beast: comparative social behaviour*. 1971.
- Kevin Langergraber, Grit Schubert, Carolyn Rowney, Richard Wrangham, Zinta Zommers, and Linda Vigilant. Genetic differentiation and the evolution of cooperation in chimpanzees and humans. *Proceedings of the Royal Society B: Biological Sciences*, 278(1717):2546–2552, 2011.
- Shishi Luo. A unifying framework reveals key properties of multilevel selection. *Journal of Theoretical Biology*, 341:41–52, 2014.
- Shishi Luo and Jonathan C Mattingly. Scaling limits of a model for selection at two scales. *Nonlinearity*, 30(4):1682, 2017.
- Gustave Malécot. *Les Mathématiques de l’Hérédité*. Masson et Cie, Paris, 1948.
- Martin A. Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.
- Martin A. Nowak, Corina E. Tarnita, and Edward O. Wilson. The evolution of eusociality. *Nature*, 466(7310):1057, 2010.
- Hisashi Ohtsuki. Does synergy rescue the evolution of cooperation? an analysis for homogeneous populations with non-overlapping generations. *Journal of Theoretical Biology*, 307:20–28, 2012.
- Hisashi Ohtsuki and Martin A. Nowak. Evolutionary games on cycles. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1598):2249–2256, 2006.
- Hisashi Ohtsuki, Christoph Hauert, Erez Lieberman, and Martin A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092):502, 2006.
- Steven Pinker. The false allure of group selection. *The handbook of evolutionary psychology*, pages 1–14, 2015.
- David C. Queller. Does population viscosity promote kin selection? *Trends in Ecology & Evolution*, 7(10):322–324, 1992.
- David C. Queller. Genetic relatedness in viscous populations. *Evolutionary Ecology*, 8(1):70–73, 1994.
- François Rousset. *Genetic structure and selection in subdivided populations*. Princeton University Press, Princeton, NJ, 2004.
- Hannes Rusch. Ancestral kinship patterns substantially reduce the negative effect of increasing group size on incentives for public goods provision. *Journal of Economic Psychology*, 64:105–115, 2018.
- Burton Simon. A dynamical model of two-level selection. *Evolutionary Ecology Research*, 12(5):555–588, 2010.
- Burton Simon, Jeffrey A. Fletcher, and Michael Doebeli. Towards a general theory of group selection. *Evolution*, 67:1561–1572, 2013.
- Elliott Sober and David Sloan Wilson. *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press, 1998.

- Peter D. Taylor. Altruism in viscous populations—an inclusive fitness model. *Evolutionary Ecology*, 6(4): 352–356, 1992a.
- Peter D. Taylor. Inclusive fitness in a homogeneous environment. In *Proc. R. Soc. Lond. B*, volume 249, pages 299–302. The Royal Society, 1992b.
- Peter D. Taylor, Troy Day, and Geoff Wild. Evolution of cooperation in a finite homogeneous graph. *Nature*, 447(7143):469–472, 2007a.
- Peter D. Taylor, Troy Day, and Geoff Wild. From inclusive fitness to fixation probability in homogeneous structured populations. *Journal of Theoretical Biology*, 249(1):101–110, 2007b.
- Arne Traulsen and Martin A. Nowak. Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, 103(29):10952–10955, 2006.
- Matthijs van Veelen. Can Hamilton’s rule be violated? *eLife*, 7:e41901, 2018.
- Matthijs van Veelen, Shishi Luo, and Burton Simon. A simple model of group selection that cannot be analyzed with inclusive fitness. *Journal of Theoretical Biology*, 360:279–289, 2014.
- Matthijs van Veelen, Benjamin Allen, Moshe Hoffman, Burton Simon, and Carl Veller. Hamilton’s rule. *Journal of Theoretical Biology*, 414:176–230, 2017.
- Robert S. Walker. Amazonian horticulturalists live in larger, more related groups than hunter–gatherers. *Evolution and Human Behavior*, 35(5):384–388, 2014.
- Bruce S. Weir and C. Clark Cockerham. Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358–1370, 1984.
- Stuart A. West, Ido Pen, and Ashleigh S. Griffin. Cooperation and competition between relatives. *Science*, 296(5565):72–75, 2002.
- David Sloan Wilson and Edward Osborne Wilson. Rethinking the theoretical foundations of socio-biology. *Quarterly Review of Biology*, 82:327–348, 2007.
- David Sloan Wilson, Gregory B. Pollock, and Lee A. Dugatkin. Can altruism evolve in purely viscous populations? *Evolutionary Ecology*, 6(4):331–341, 1992.

Supplementary Information

S.1 Introduction

S.1.1 The cancellation effect at the individual level

The cancellation effect at the individual level was discovered by Wilson et al. (1992) and Taylor (1992a;b). Before then, it was more or less generally thought that as soon as interacting individuals are related, there is scope for cooperation to evolve (see Hamilton, 1971; Boyd, 1982; Grafen, 1983; and other references in Wilson et al., 1992; Taylor, 1992a; and Taylor, 1992b, for exceptions). The idea was that positive relatedness results in cooperators being more likely to interact with cooperators compared to defectors, which implies that, while cooperators pay the cost of cooperating, they will also be on the receiving end of cooperation more often. If they are indeed sufficiently much more often the recipients of cooperation than defectors are, and if the benefits are sufficiently large, then the cost of cooperation can be offset by the increase in benefits received. The idea was that positive relatedness means that cooperators are more likely to interact with cooperators than defectors are, which implies that, while cooperators pay the cost of cooperating, they will also be on the receiving end of cooperation more often. If they are indeed sufficiently much more often the recipients of cooperation than defectors are, and if the benefits are sufficiently large, then the cost of cooperation can be offset by the increase in benefits received.

What Wilson et al. (1992) and Taylor (1992a;b) discovered is that being around cooperators is not necessarily unambiguously good news. While being around more cooperators means receiving more cooperation – which is good – it can also mean being around individuals that cooperate more with one another, and therefore constitute more fierce competition – which is not good. For cases in which relatedness is caused by local dispersal, receiving more cooperation and facing more intense competition can go hand in hand, and therefore, if individuals have opportunities for cooperating that are as local as their competition is, no benefit is large enough to get costly cooperation to evolve. The main insight provided by Wilson et al. (1992) and Taylor (1992a;b) is therefore that it is not enough to be related; what is needed is a discrepancy between the relatednesses to the individuals with whom one has the opportunity to cooperate, and those with whom one has to compete. While the idea that positive relatedness alone would be enough for the evolution of cooperation was inspired by Hamilton’s rule (Hamilton, 1964a;b), it was pointed out that, by defining the fitness effects of cooperation versus defection appropriately, the cancellation effect can actually be identified within the framework of Hamilton’s rule; see Taylor (1992b); Grafen (2007); Allen et al. (2012), and Section 7 in van Veelen et al. (2017), all in settings where the game between individuals satisfies equal gains from switching.¹

S.1.2 The cancellation effect at the group level

Group selection models aim at capturing the opposing effects of selection at the individual level, where defectors do better than cooperators within groups, and selection at the group level, where groups with more cooperators do better than groups with fewer cooperators (Sober and Wilson, 1998; Wilson and Wilson,

¹Ohtsuki (2012) analyzes a model that does not satisfy equal gains from switching. Other papers on the scale of cooperation versus the scale of competition are Queller (1992; 1994) and West et al. (2002), but these papers take a different approach. While the other papers include the cancellation effect by making sure to account for all fitness effects, and keeping the definition of relatedness the same, the papers in this set get Hamilton’s rule to hold by changing the relatedness into *effective relatedness*. When we will do computations to include the cancellation effect at the group level, we will follow the first approach.

2007). The existing models within the group selection literature share the property that competition between individuals happens within groups, and that competition between groups happens in a setting where all groups compete with all other groups equally intensely (Traulsen and Nowak, 2006; Simon, 2010; Simon et al., 2013; Luo, 2014; Luo and Mattingly, 2017; van Veelen et al., 2014). This last property is a useful simplification if the aim is to illustrate the possibility of a tug of war between the different levels of selection. It is, however, not particularly realistic. Groups themselves typically live in a structured population of groups, where they compete with their neighbouring groups more often than they do with groups that are farther apart. Local dispersal would then imply that groups with many cooperators are typically surrounded by groups that also contain many cooperators, compared to the groups that surround groups with many defectors, and therefore are also subject to more intense competition. This can significantly dampen the benefits of being a cooperative group, which in turn affects the balance between selection at the individual and at the group level.

In order to study the cancellation effect at the group level, and how it affects the balance between selection at different levels, we consider a stylized model, in which groups live on a cycle. We look at two replacement rules for groups. The first replacement rule is Birth-Death, where groups replace their direct neighbours. The second replacement rule is Shift, where a group at one position can reproduce, and a group anywhere else can die, and all groups in between the two just move over. With Birth-Death, groups compete with their direct neighbours, and the cancellation effect at the group level is the largest it can be. With Shift, there is no cancellation at the group level at all.

The rest of the Supplementary Information is organized as follows. In Section S.2, we describe the model, including both replacement rules. Analytical thresholds for the benefit-to-cost ratios in the limit of weak selection, both for Birth-Death and Shift, are derived in Section S.3. These thresholds are expressed in terms of the parameters of the model and relatednesses. The relatednesses are endogenous themselves and also depend on the parameters of the model, so we compute these in Section S.4. In Section S.5, the benefit-to-cost ratios and the relatednesses are combined and it is shown that the critical benefit-to-cost ratio for the Birth-Death process is always higher than the critical benefit-to-cost ratio for the Shift process, making group selection models that assume away the cancellation effect more optimistic about the conditions for the evolution of cooperation. These thresholds are compared to simulation results, not in the limit of weak selection, in Section S.6. In Section S.7, we derive analytical results for some limits other than the limit of weak selection, and in Section S.8 we discuss the empirical implications.

S.2 The model

S.2.1 Two update rules on the cycle

We consider a simple model, in which groups are situated on a cycle. Both the group size and the number of groups are fixed; at each point in time, there are m groups consisting of n individuals. Each individual can be either a cooperator (C) or a defector (D). In every time period, one of three types of events will happen: an individual can replace another individual within a group, a group can replace another group, or two individuals from two neighbouring groups can change places. These individual, group, and migration events happen with probabilities p , q and r , respectively, and, without loss of generality, we assume that $p + q + r = 1$.

In order to demonstrate the difference between global and local between group competition, we compare two different replacement rules for group reproduction: Birth-Death (BD) and Shift. Competition between groups is local in BD, and global in Shift. Individual and migration events happen in the same way in both processes.

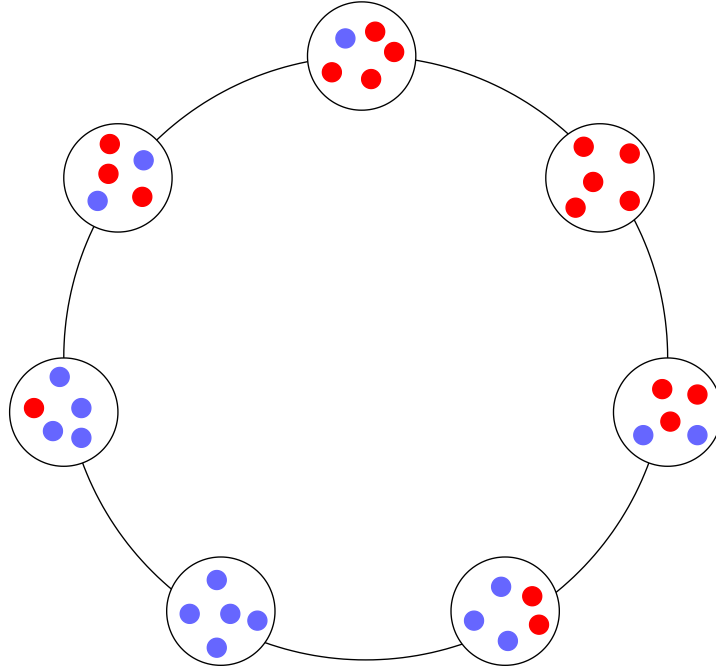


Figure S.1: An example of a population state on a cycle with $m = 7$ groups of $n = 5$ individuals each. The blue dots indicate cooperators and the red dots indicate defectors. Each group has one right and one left neighbour.

S.2.1.1 Individual events

If an individual event occurs – which happens with probability p – then a random group is selected, and within that group, an individual is chosen to produce an identical offspring. All groups have equal probability of being chosen to host an individual level event. Within the group, defectors get an individual payoff of 1 and cooperators get an individual payoff of $1 - c$. The intensity of selection at the individual level w_{ind} is

then used to transform these payoffs to values f_C and f_D :

$$f_C = 1 - w_{ind} + w_{ind}(1 - c) = 1 - w_{ind}c \quad \text{and} \quad f_D = 1 - w_{ind} + w_{ind}(1) = 1$$

The probabilities with which individuals are chosen for reproduction within the group are proportional to these values.² The probability $p_C(k_i)$ that a cooperator is chosen, and the probability $p_D(k_i)$ that a defector is chosen, in a group with k_i cooperators, then become:

$$p_C(k_i) = \frac{k_i f_C}{k_i f_C + (n - k_i) f_D} = \frac{k_i(1 - w_{ind}c)}{n - k_i w_{ind}c}$$

$$p_D(k_i) = \frac{(n - k_i) f_D}{k_i f_C + (n - k_i) f_D} = \frac{n - k_i}{n - k_i w_{ind}c} = 1 - p_C(k_i)$$

Whenever an individual reproduces, someone from the same group is chosen to die, where all individuals, including the parent, but excluding the offspring, are chosen with probability $\frac{1}{n}$.

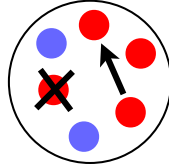


Figure S.2: An example of the individual reproduction events. A defector is chosen for reproduction and produces an identical offspring. Another defector is chosen to die. In this case, the overall group composition has not changed.

If an individual event happens, the number of cooperators in that group can go up or down by one, or remain constant, depending on who reproduces and who dies within the group. Once a group reaches a state where all individuals are cooperators or all individuals are defectors, individual selection can not change the state of that group. Individual selection on its own, within a given group, therefore constitutes a Markov process with two absorbing states.

S.2.1.2 Group events

If a group event occurs – which happens with probability q – then one group is chosen to reproduce, and one group is chosen to die. Which group reproduces depends on the distribution of cooperators among groups. If the groups are numbered from 1 to m , then a population state is a vector k , in which $k_i \in \{0, 1, \dots, n\}$ is the number of cooperators in group i , for $i = 1, \dots, m$. Groups i and $i + 1$ are neighbouring groups, for $i = 1, \dots, m - 1$, as well as groups 1 and m , which makes this a cycle. The group payoff of group i is $1 + \frac{k_i}{n}b$. The intensity of selection at the group level w_{gr} is then used to transform these payoffs to values

$$g_i(k_i) = 1 - w_{gr} + w_{gr} \left(1 + \frac{k_i}{n}b \right) = 1 + w_{gr} \frac{k_i}{n}b$$

²One could call these values "fitnesses", in line with Nowak (2006). We chose to not use that term here, and reserve the word fitness, and fitness effects, for (effects on) expected numbers of offspring.

Both in BD and in Shift, first a group is chosen for reproduction, where each group's probability of being chosen is proportional to the values $g_i(k_i)$ as given below.

$$q_i(k_i, K) = \frac{g_i(k_i)}{\sum_{j=1}^m g_j(k_j)} = \frac{1 + w_{gr} \frac{k_i}{n} b}{n + w_{gr} \frac{K}{n} b}$$

where $K = \sum_{j=1}^m k_j$ is the total number of cooperators in the population. If a group is chosen for reproduction, it produces an identical offspring group. Which group is being replaced depends on which replacement rule is used at the group level.

Birth-Death (BD)

The offspring group replaces its own parent group with probability $\frac{1}{m}$, and it replaces either the left or the right neighbour of the parent group, both with probability $\frac{m-1}{2m}$. If the offspring group replaces the parent group, the population state does not change. The possibility to replace the parent group is included in order to make the analytical comparison between the two processes more straightforward. This does not have any profound consequences; any process with that possibility is equivalent to a process without that possibility, and with a lower probability of having a group reproduction event at all.

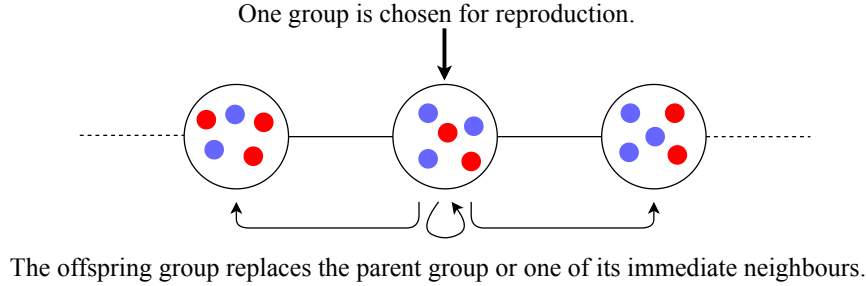


Figure S.3: The BD process. One of the groups is chosen to reproduce, proportional to group values, and produces an identical offspring group. The offspring group replaces one of the neighbouring groups, or, with a small probability, it replaces the parent group itself.

Shift

All groups, including the parent group, but excluding the offspring group, die with probability $\frac{1}{m}$. Hence, with Shift, every group is equally likely to die and competition between groups is therefore global, as it is in the standard group selection models with a well-mixed population of groups. Once a group dies, the offspring group is either placed to the right or to the left of the parent group, with equal probability, and every other group between the parent group and the dying group moves over one spot.

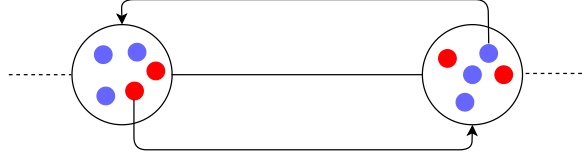


Figure S.5: An example of a migration event, where a defector from the group on the left and a cooperator from the group on the right change places

S.2.2 Alternative modeling choices

There are places in the model where, with the literature on group selection in mind, one can easily think of other ways to determine who reproduces and who is replaced. The reason for the choices we made is typically that they allow us to derive relatively tractable analytical solutions, which help illustrate the cancellation effect. Alternative choices would produce the same gap between full cancellation at the group level (BD) and no cancellation at the group level (Shift), but some would come with (much) more complicated analytical solutions.

Luo (2014)

In Luo (2014) and van Veelen et al. (2014), the reproduction rate of an individual is an individual characteristic, where defectors have a higher individual reproduction rate than cooperators do. That implies that a group with more defectors is more likely to host an individual reproduction event than a group with fewer defectors. In our model, every group is equally likely to host an individual reproduction event. Another implication of the choices in Luo (2014) and van Veelen et al. (2014) is that there, the ratio of group level reproduction events to individual level reproduction events depends on how many cooperators there are in the population as a whole; with many cooperators, group level reproduction events happen more frequently, relative to individual level reproduction events, compared to a population with fewer cooperators. In our model, this ratio is $\frac{p}{q}$, which is constant.

The modeling choices in Luo (2014), and similar ones in Simon (2010) and Simon et al. (2013), produce a model that is in some respects more elegant than ours. When combined with a structured population of groups like the cycle, and with different replacement rules for group reproduction, the cancellation effect would be present and absent in their model in the same way as it is in our version. While our model may be a bit less elegant, it does allow for a more straightforward derivation of the formulas for the critical b/c ratios. (Just to be sure: groups in Luo (2014) and van Veelen et al. (2014) compete globally, making this otherwise equivalent to the Shift model without migration).

A final difference, not in the model itself, but in the approach to deriving analytical solutions, is that Simon (2010); Simon et al. (2013); Luo (2014), and van Veelen et al. (2014) all take limits of group size n and number of groups m going to infinity, while we derive analytical solutions in the limit of weak selection.

Traulsen & Nowak (2006)

In Traulsen and Nowak (2006), individuals within groups affect their own and each other's individual reproduction rates. Cooperators lower their own individual reproduction rate, and increase the individual reproduction rates of their fellow group members. When an individual reproduction event happens, there is a small probability that, instead of the offspring replacing another group member, the group splits. Because

individuals in all cooperator groups reproduce more frequently than individuals in all defector groups, all cooperator groups also split more often, and therefore produce more offspring groups. This is a bit further removed from our model, but it would of course be possible to make versions of this model where groups are situated on a cycle, and, in case a group splits, it either replaces a neighbouring group (BD), or a randomly chosen group (Shift). We would also have to include (local) migration, because, similar to our setting, we would expect the difference between the two replacement rules to disappear in the limit of no migration. Traulsen and Nowak (2006) do have a version with global migration, but the method for finding analytical solutions is not directly adaptable to local migration. We find that in our model, the number of groups m , group size n , and migration rate r all affect the critical b/c ratio in ways similar to how they affect that ratio in their model. Of course the cancellation effect cannot be identified in their model, unless a structured population of groups is introduced.

Public goods game

Although the game in our model does have the key properties of a public goods game, where paying individual costs come with collective benefits, one can also define payoffs so that the game looks like a public goods game already at the individual level. Such a version of the model would leave all model assumptions unchanged, except for the individual and group payoff functions. The values for a cooperator and a defector in group i , in which there are k_i cooperators, then become:

$$f_C^{PGG}(k_i) = 1 - w_{ind} + w_{ind} \left(1 + \frac{k_i}{n}b - c \right) = 1 + w_{ind} \left(\frac{k_i}{n}b - c \right)$$

and

$$f_D^{PGG}(k_i) = 1 - w_{ind} + w_{ind} \left(1 + \frac{k_i}{n}b \right) = 1 + w_{ind} \left(\frac{k_i}{n}b \right)$$

where w_{ind} is the intensity of selection at the individual level. The value for the group is defined as:

$$g_i(k_i)^{PGG} = 1 - w_{gr} + w_{gr} \left(1 + \frac{k_i}{n}(b - c) \right) = 1 + w_{gr} \left(\frac{k_i}{n}(b - c) \right)$$

where w_{gr} is the intensity of selection at the group level. We assume that $b > c$.

In this formulation, the payoffs seem to make this a public goods game, already at the individual level, as all individuals get a higher payoff when all individuals play C (when they all get $b - c$), compared to what all individuals get when all play D (when they all get 0). In our setting, this is however not really an improvement for all at the individual level. The probability of a group being chosen to host an individual event is $\frac{1}{m}$, and this probability is therefore independent of the population state. When all individuals have the same payoff, whether it is all large or all small, their probability of reproducing, conditional on their group being chosen to host the individual reproduction event, is $\frac{1}{n}$. These payoffs therefore only seemingly introduce the public goods nature at the individual level. In alternative settings, where groups with many cooperators also host more individual reproduction events than groups with many defectors, the reproduction rates of everyone will be higher if everyone is a cooperator rather than everyone being a defector, but now this is compensated by an also elevated death rate, again neutralizing all “gains from cooperation” at the individual reproduction level. The only difference between the formulations is how the change in individual reproduction rate depends on the total number of cooperators in the group. In the formulation we chose, being a cooperator and not a defector decreases ones individual reproduction rate by $\frac{1}{n-k_ic} - \frac{1-c}{n-(k_i+1)c}$,

if k_i is the number of cooperators among the other members of the group. In this “PGG” formulation, that difference is $\frac{1+\frac{k_i}{n}b}{n+k_i(b-c)} - \frac{1+\frac{k_i+1}{n}b-c}{n+(k_i+1)(b-c)}$. In other words, in our formulation, the reduction in individual reproduction rate gets a bit larger when more others cooperate; while in this alternative PGG formulation, the reduction in individual reproduction rate gets a bit smaller.

This alternative formulation does furthermore link individual and group values, by making the latter the average of the former. The gains in group reproduction rate relate in a more straightforward way; the $b - c$ in the PGG version replaces the b in our version.

Given that the effective differences between these two versions of the model are only minor details, it is not surprising that simulations also give very similar results. For obtaining analytic results, our version is easier to work with.

S.3 Analytical results in the limit of weak selection

We would like to derive critical b/c ratios, above which cooperation is selected for, in the limit of weak selection. In order to do that, we will go over the effects of being a cooperator instead of a defector on reproduction rates and death rates. The b and the c are just model parameters, so they are not the fitness benefits and fitness costs of cooperation. The effects that we compute below do amount to those fitness benefits and costs. Because these effects satisfy equal gains from switching locally (they are additive in the limit of weak selection), we can follow an inclusive fitness approach, where these effects are weighted with the relatednesses to the individual that the effects are on, in order to determine the direction of selection (van Veelen et al., 2017; van Veelen, 2018). We will use this approach to derive critical b/c ratios, which are therefore formulated in terms of model parameters. In the following subsections, we consider a case where one individual switches from being a defector to being a cooperator, and we calculate the effects of this change by the focal individual on every individual in the population, weighted by the corresponding relatednesses.

Since the probabilities concerning reproduction events, both at the individual level and at the group level, are the same in BD and in Shift, the effects of being a cooperator instead of a defector on reproduction rates will be the same for both. Moreover, in both processes, the individual death rate is the same. The difference between the two processes, therefore, will only start showing up when we compute the effects on group death rates.

S.3.1 Changes in reproduction rates

S.3.1.1 Changes in individual reproduction rates

Changes in my reproduction rate

If I am a C instead of a D , with i other C players in my group, and my group is chosen for an individual update, I change my probability of being chosen for reproduction from $\frac{1}{n-iw_{ind}c}$ to $\frac{1-w_{ind}c}{n-(i+1)w_{ind}c}$, which is a difference of $\frac{1-w_{ind}c}{n-(i+1)w_{ind}c} - \frac{1}{n-iw_{ind}c}$. If we take the derivative with respect to w_{ind} for both terms, we get:

$$\frac{d\left(\frac{1-w_{ind}c}{n-(i+1)w_{ind}c}\right)}{dw_{ind}} = -\frac{n-(i+1)}{(n-(i+1)w_{ind}c)^2} \cdot c \quad \text{and} \quad \frac{d\left(\frac{1}{n-iw_{ind}c}\right)}{dw_{ind}} = \frac{i}{(n-iw_{ind}c)^2} \cdot c$$

Evaluated at $w_{ind} = 0$, we get:

$$\left.\frac{d\left(\frac{1-w_{ind}c}{n-(i+1)w_{ind}c}\right)}{dw_{ind}}\right|_{w_{ind}=0} = -\frac{n-(i+1)}{n^2} \cdot c \quad \text{and} \quad \left.\frac{d\left(\frac{1}{n-iw_{ind}c}\right)}{dw_{ind}}\right|_{w_{ind}=0} = \frac{i}{n^2} \cdot c$$

which means that the change in probability of being chosen for reproduction, for w_{ind} close to 0 (weak selection), can be approximated by

$$\frac{1-w_{ind}c}{n-(i+1)w_{ind}c} - \frac{1}{n-iw_{ind}c} \approx -\frac{n-1}{n^2}c \cdot w_{ind}$$

Changes in the reproduction rate of a C in my group

If I am a C instead of a D , with i other C players in my group, and my group is chosen for an individual update, I change other C 's individual probability of being chosen for reproduction from $\frac{1-w_{ind}c}{n-iw_{ind}c}$ to $\frac{1-w_{ind}c}{n-(i+1)w_{ind}c}$, which is a difference of $\frac{1-w_{ind}c}{n-(i+1)w_{ind}c} - \frac{1-w_{ind}c}{n-iw_{ind}c}$. If we take the derivative with respect to w_{ind} , we get:

$$\frac{d\left(\frac{1-w_{ind}c}{n-(i+1)w_{ind}c}\right)}{dw_{ind}} = -\frac{n-(i+1)}{(n-(i+1)w_{ind}c)^2} \cdot c \quad \text{and} \quad \frac{d\left(\frac{1-w_{ind}c}{n-iw_{ind}c}\right)}{dw_{ind}} = -\frac{n-i}{(n-iw_{ind}c)^2} \cdot c$$

Evaluated at $w_{ind} = 0$, we get:

$$\left.\frac{d\left(\frac{1-w_{ind}c}{n-(i+1)w_{ind}c}\right)}{dw_{ind}}\right|_{w_{ind}=0} = -\frac{n-(i+1)}{n^2} \cdot c \quad \text{and} \quad \left.\frac{d\left(\frac{1-w_{ind}c}{n-iw_{ind}c}\right)}{dw_{ind}}\right|_{w_{ind}=0} = -\frac{n-i}{n^2} \cdot c$$

which means that the change in their probability of being chosen for reproduction, for w_{ind} close to 0 (weak selection), can be approximated by

$$\frac{1-w_{ind}c}{n-(i+1)w_{ind}c} - \frac{1-w_{ind}c}{n-iw_{ind}c} \approx \frac{1}{n^2}c \cdot w_{ind}$$

Changes in the reproduction rate of a D in my group

If I am a C instead of a D , with i other C players in my group, and my group is chosen for an individual update, I change other D 's individual chance of being chosen for reproduction from $\frac{1}{n-iw_{ind}c}$ to $\frac{1}{n-(i+1)w_{ind}c}$, which is a difference of $\frac{1}{n-(i+1)w_{ind}c} - \frac{1}{n-iw_{ind}c}$. If we take the derivative with respect to w_{ind} , we get:

$$\frac{d\left(\frac{1}{n-(i+1)w_{ind}c}\right)}{dw_{ind}} = \frac{i+1}{(n-(i+1)w_{ind}c)^2} \cdot c \quad \text{and} \quad \frac{d\left(\frac{1}{n-iw_{ind}c}\right)}{dw_{ind}} = \frac{i}{(n-iw_{ind}c)^2} \cdot c$$

Evaluated at $w_{ind} = 0$, we get:

$$\left.\frac{d\left(\frac{1}{n-(i+1)w_{ind}c}\right)}{dw_{ind}}\right|_{w_{ind}=0} = \frac{i+1}{n^2} \cdot c \quad \text{and} \quad \left.\frac{d\left(\frac{1}{n-iw_{ind}c}\right)}{dw_{ind}}\right|_{w_{ind}=0} = \frac{i}{n^2} \cdot c$$

which means that the change in their probability of being chosen for reproduction, for w_{ind} close to 0 (weak selection), can be approximated by

$$\frac{1}{n-(i+1)w_{ind}c} - \frac{1}{n-iw_{ind}c} \approx \frac{1}{n^2}c \cdot w_{ind}$$

The effects on individual reproduction rates should be multiplied by $p_m^{\frac{1}{m}}$ in order to account for the probability with which an individual event happens, and that my group is chosen to host it. The changes in individual reproduction rates within the group add up to 0, as they should with a fixed group size. There are no effects on individual reproduction rates in other groups.

S.3.1.2 Changes in group reproduction rates

Changes in my group's reproduction rate

If I am a C instead of a D , with i other C players in my group, $i = 0, 1, \dots, n-1$, and j other C 's in the population as a whole, $j = i, i+1, \dots, i+n(m-1)$, then if a group level event happens, I change my group's probability of being chosen for reproduction from $\frac{1+w_{gr}\frac{i}{n}b}{m+w_{gr}\frac{j}{n}b}$ to $\frac{1+w_{gr}\frac{i+1}{n}b}{m+w_{gr}\frac{j+1}{n}b}$, which is a difference of $\frac{1+w_{gr}\frac{i+1}{n}b}{m+w_{gr}\frac{j+1}{n}b} - \frac{1+w_{gr}\frac{i}{n}b}{m+w_{gr}\frac{j}{n}b}$. If we take the derivative of both terms with respect to w_{gr} , we get:

$$\frac{d\left(\frac{1+w_{gr}\frac{i+1}{n}b}{m+w_{gr}\frac{j+1}{n}b}\right)}{dw_{gr}} = \frac{\frac{m(i+1)-(j+1)}{n}}{(m+w_{gr}\frac{j+1}{n}b)^2} \cdot b \quad \text{and} \quad \frac{d\left(\frac{1+w_{gr}\frac{i}{n}b}{m+w_{gr}\frac{j}{n}b}\right)}{dw_{gr}} = \frac{\frac{mi-j}{n}}{(m+w_{gr}\frac{j}{n}b)^2} \cdot b$$

Evaluated at $w_{gr} = 0$, we get:

$$\left.\frac{d\left(\frac{1+w_{gr}\frac{i+1}{n}b}{m+w_{gr}\frac{j+1}{n}b}\right)}{dw_{gr}}\right|_{w_{gr}=0} = \frac{m(i+1)-(j+1)}{nm^2} \cdot b \quad \text{and} \quad \left.\frac{d\left(\frac{1+w_{gr}\frac{i}{n}b}{m+w_{gr}\frac{j}{n}b}\right)}{dw_{gr}}\right|_{w_{gr}=0} = \frac{mi-j}{nm^2} \cdot b$$

which means that the change in my group's probability of being chosen for reproduction, for w_{gr} close to 0 (weak selection), can be approximated by

$$\frac{1+w_{gr}\frac{i+1}{n}b}{m+w_{gr}\frac{j+1}{n}b} - \frac{1+w_{gr}\frac{i}{n}b}{m+w_{gr}\frac{j}{n}b} \approx \frac{m-1}{nm^2}b \cdot w_{gr}$$

Changes in other groups' reproduction rates

If I am a C instead of a D , with i other C players in my group, $i = 0, 1, \dots, n-1$, and j other C 's in the population as a whole, $j = i, i+1, \dots, i+n(m-1)$, then if a group level event happens, I change the probability of being chosen for reproduction of a random other group with k cooperators from $\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j}{n}b}$ to $\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j+1}{n}b}$, which is a difference of $\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j+1}{n}b} - \frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j}{n}b}$. If we take the derivative with respect to w_{gr} , we get:

$$\frac{d\left(\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j+1}{n}b}\right)}{dw_{gr}} = \frac{\frac{mk-(j+1)}{n}}{(m+w_{gr}\frac{j+1}{n}b)^2} \cdot b \quad \text{and} \quad \frac{d\left(\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j}{n}b}\right)}{dw_{gr}} = \frac{\frac{mk-j}{n}}{(m+w_{gr}\frac{j}{n}b)^2} \cdot b$$

Evaluated at $w_{gr} = 0$, we get:

$$\left.\frac{d\left(\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j+1}{n}b}\right)}{dw_{gr}}\right|_{w_{gr}=0} = \frac{mk-(j+1)}{nm^2} \cdot b \quad \text{and} \quad \left.\frac{d\left(\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j}{n}b}\right)}{dw_{gr}}\right|_{w_{gr}=0} = \frac{mk-j}{nm^2} \cdot b$$

which means that the change in that other group's probability of being chosen for reproduction, for w_{gr} close to 0 (weak selection), can be approximated by

$$\frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j+1}{n}b} - \frac{1+w_{gr}\frac{k}{n}b}{m+w_{gr}\frac{j}{n}b} \approx -\frac{1}{nm^2}b \cdot w_{gr}$$

The effects on group reproduction rates should be multiplied by q in order to account for the probability with which a group event happens. The changes in group reproduction rates add up to 0, as they should with a fixed number of groups.

S.3.1.3 Overall effect through changes in reproduction rates

The combined effects, close to neutrality, all weighted with the corresponding relatednesses, are:

$$\frac{p}{m} \left(-\frac{n-1}{n^2} + (n-1)r_s \frac{1}{n^2} \right) c \cdot w_{ind} + q \left(\frac{m-1}{nm^2} + (n-1)r_s \frac{m-1}{nm^2} - n(m-1)r_o \frac{1}{nm^2} \right) b \cdot w_{gr}$$

where r_s is the relatedness between an individual and a randomly chosen other individual from the same group, and r_o is the relatedness between an individual and a randomly chosen individual from a randomly chosen other group.

In Section S.4, we derive identities concerning different relatednesses. We can use one of those to rewrite the combined effects. Equation (37) states that $r_o = -\frac{1}{n(m-1)} - \frac{n-1}{n(m-1)}r_s$, and using that, we can see rewrite the combined effects as

$$\begin{aligned} \frac{p}{m} \left(-\frac{n-1}{n^2} + (n-1)r_s \frac{1}{n^2} \right) c \cdot w_{ind} + q \left(\frac{m-1}{nm^2} (1 + (n-1)r_s) + (1 + (n-1)r_s) \frac{1}{nm^2} \right) b \cdot w_{gr} \\ = -p \frac{n-1}{mn^2} (1 - r_s) c \cdot w_{ind} + q \frac{1}{nm} (1 + (n-1)r_s) b \cdot w_{gr} \end{aligned}$$

S.3.2 Changes in death rates

S.3.2.1 Changes in individual death rates

Individual death rates do not change when a player switches from being D to C .

S.3.2.2 Changes in group death rates

Changes in group death rates differ between the two update processes at the group level.

Birth-Death

Changes in my group's death rate

If I am a C instead of a D , with i other C players in my group, $i = 0, 1, \dots, n-1$, and j other C 's in the population as a whole, $j = i, i+1, \dots, i+n(m-1)$, then, if a group level event happens, we have seen that I change the reproduction rate of my own group by approximately $\frac{m-1}{nm^2} b \cdot w_{gr}$. Since with probability $\frac{1}{m}$ the offspring group replaces the parent group, that comes with an increase in death rate equal to

$$\frac{1}{m} \frac{m-1}{nm^2} b \cdot w_{gr}$$

We have also seen that I change the reproduction rate of all other groups by approximately $-\frac{1}{nm^2} b \cdot w_{gr}$, including my two neighbouring groups. If one of those is chosen for reproduction, it replaces my group with

probability $\frac{m-1}{2m}$, reducing my death rate by

$$2 \cdot \frac{1}{2} \frac{m-1}{m} \frac{1}{nm^2} b \cdot w_{gr}$$

These two cancel out exactly, so the overall effect on the death rate of my group is 0, as my neighbours are now less likely to be chosen for reproduction and replace my group, but my group is more likely to reproduce and replace itself.

Changes in the death rate of the two neighbouring groups

Following a similar argument, we find that the probability that one of my next-door neighbour groups is replaced changes by:

$$\frac{1}{2} \frac{m-1}{m} \frac{m-1}{nm^2} b \cdot w_{gr} - \frac{1}{2} \frac{m-1}{m} \frac{1}{nm^2} b \cdot w_{gr} - \frac{1}{m} \frac{1}{nm^2} b \cdot w_{gr}$$

The first term is the product of the probability that, if my group is chosen for reproduction, it replaces a given neighbour, and the increase in my groups probability to reproduce. The second term is the product of the probability of the neighbouring group of the neighbouring group to replace the neighbouring group, if chosen for reproduction, and the decrease in their probability of being chosen for reproduction. The third term is the product of the probability of the neighbouring group to replace itself, if chosen for reproduction, and the decrease in their probability of being chosen for reproduction. This sum can be rewritten as:

$$\left(\frac{1}{2} (m-1)(m-2) - 1 \right) \frac{1}{nm^3} b \cdot w_{gr} = \frac{1}{2} (m-3) \frac{1}{nm^2} b \cdot w_{gr}$$

This effect is the same for both right and left next-door neighbours.

Changes in the death rates of other groups

The probability of being replaced changes by:

$$2 \cdot -\frac{1}{2} \frac{m-1}{m} \frac{1}{nm^2} b \cdot w_{gr} - \frac{1}{m} \frac{1}{nm^2} b \cdot w_{gr} = -\frac{1}{nm^2} b \cdot w_{gr}$$

for any of the $m-3$ other groups. The first term on the left hand side is twice the product of the probability that a given neighbouring group replaces a given group, times the change in reproduction probability of those neighbouring groups. The second term is the product of the probability a group replaces its parent group, when chosen for reproduction, and the change in reproduction probability of such a group.

The effects on group reproduction rates should be multiplied by q in order to account for the probability with which a group event happens.

Shift

Changes in group death rates

For the Shift process, group death rates do not change when a player switches from being D to C .

S.3.2.3 Overall effect through changes in death rates

Birth-Death

The overall effect through death rates again combines the effects on individual death rates (which are zero) and the effects on group death rates affecting self, group members, and members of other groups, each weighted with the corresponding relatedness measure:

$$\begin{aligned}
& p \frac{1}{m} \cdot 0 + q \left(0 - (n-1)r_s \cdot 0 + nr_1 \frac{1}{2} \frac{m-3}{nm^2} b \cdot w_{gr} - n \left(\sum_{i=2}^{m-2} r_i \right) \frac{1}{nm^2} b \cdot w_{gr} + nr_{m-1} \frac{1}{2} \frac{m-3}{nm^2} b \cdot w_{gr} \right) \\
&= q \left(nr_1 \frac{1}{2} \frac{m-3}{nm^2} + nr_1 \frac{1}{nm^2} - n \left(\sum_{i=1}^{m-1} r_i \right) \frac{1}{nm^2} + nr_{m-1} \frac{1}{nm^2} + nr_{m-1} \frac{1}{2} \frac{m-3}{nm^2} \right) b \cdot w_{gr} \\
&= q \left(nr_1 \frac{m-3}{nm^2} + 2nr_1 \frac{1}{nm^2} - n \left(\sum_{i=1}^{m-1} r_i \right) \frac{1}{nm^2} \right) b \cdot w_{gr} \\
&= q \left(nr_1 \frac{m-1}{nm^2} + (1 + (n-1)r_s) \frac{1}{nm^2} \right) b \cdot w_{gr} \\
&= q \frac{1}{nm^2} (n(m-1)r_1 + 1 + (n-1)r_s) b \cdot w_{gr}
\end{aligned}$$

where we used the identity $r_1 = r_{m-1}$ and Equation (35).

Shift

For Shift, the overall effect through changes in death rates is zero, since the effects on death rates of both individuals and groups are zero.

S.3.3 Overall effect of switching from D to C

The overall effect of a player switching from playing D to C would be the effect on reproduction rates minus the effect on death rates.

S.3.3.1 Birth-Death

The overall effect for the BD process is

$$\begin{aligned}
& -p \frac{n-1}{mn^2} (1-r_s)c \cdot w_{ind} + q \frac{m}{nm^2} (1 + (n-1)r_s) b \cdot w_{gr} - q \frac{1}{nm^2} (n(m-1)r_1 + 1 + (n-1)r_s) b \cdot w_{gr} \\
&= -p \frac{n-1}{mn^2} (1-r_s)c \cdot w_{ind} + q \frac{1}{nm^2} ((m-1)(1 + (n-1)r_s) - (m-1)nr_1) b \cdot w_{gr} \\
&= -p \frac{n-1}{mn^2} (1-r_s)c \cdot w_{ind} + q \frac{m-1}{nm^2} (1 + (n-1)r_s - nr_1) b \cdot w_{gr}
\end{aligned}$$

For $w_{ind} = w_{gr}$, this effect is positive if

$$\begin{aligned}
& -p \frac{n-1}{mn^2} (1-r_s)c + q \frac{m-1}{nm^2} (1 + (n-1)r_s - nr_1) b > 0 \\
& q \frac{m-1}{nm^2} (1 + (n-1)r_s - nr_1) b > p \frac{n-1}{mn^2} (1-r_s)c
\end{aligned}$$

which implies that, in the BD process, cooperation is selected for if

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{m}{m-1} \frac{1-r_s}{1+(n-1)r_s - nr_1} \quad (3)$$

S.3.3.2 Shift

Since the effect on death rates is zero, the overall effect would be equal to the effect on reproduction rates:

$$-p \frac{n-1}{mn^2} (1-r_s)c \cdot w_{ind} + q \frac{m}{nm^2} (1+(n-1)r_s) b \cdot w_{gr}$$

For $w_i = w_g$, this effect is positive if

$$\begin{aligned} -p \frac{n-1}{mn^2} (1-r_s)c + q \frac{m}{nm^2} (1+(n-1)r_s) b &> 0 \\ q \frac{m}{nm^2} (1+(n-1)r_s) b &> p \frac{n-1}{mn^2} (1-r_s)c \end{aligned}$$

which implies that, in the Shift process, cooperation is selected for if

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{1-r_s}{1+(n-1)r_s} \quad (4)$$

S.3.4 Comparing thresholds

There are three differences between these two thresholds. The first is that Equation (3) has an $\frac{m}{m-1}$ term that is absent in Equation (4). For a large number of groups m , this term approaches 1, which means that for the cases we are most interested in, this does not contribute too much to the gap between the two b/c ratios. The second difference is that r_s will not be the same across the two processes, even if everything else (that is: p , q , r , n and m) is equal. In Section S.4, we calculate how r_s depends on those five parameters for both processes, and it turns out that r_s is higher for BD than for Shift. Therefore, if it was not for the third difference, the critical b/c ratio would actually be *lower* for BD than for Shift. The third difference is that in the last fraction in the equation, there is a $-nr_1$ term in the denominator for BD, but not for Shift. With BD, groups replace their direct neighbours, and for the cancellation effect it is important how similar these direct neighbours are. That is why r_1 appears in the formula for BD, and not in the formula for Shift, where groups compete globally, and the cancellation effect is absent. In Sections S.5 and S.6, we will see that this makes the threshold for BD end up higher than that for Shift.

S.3.5 Hamilton's rule

We derived these thresholds using inclusive fitness. Therefore, it is worth pointing out that the b and the c in those thresholds are model parameters, and that they are *not* the fitness benefits and fitness costs of cooperation. It would therefore not be correct to read these formulas themselves as versions of Hamilton's rule, where the b and the c would represent the fitness benefits and costs, and the right hand side would replace $\frac{1}{r}$ (Ohtsuki et al., 2006; Nowak et al., 2010). These parameters do however determine the size of the fitness effects, as we have seen in the derivations that preceded Equation (3) and (4).

S.4 Relatedness

In this section, we will calculate relatednesses r_i between individuals that are i groups apart, both for Birth-Death and for Shift, in the limit of weak selection, using identical-by-descent (IBD) probabilities. Two individuals are considered IBD if they descend from a common ancestor, and no mutations have occurred along their lineage. We will first derive the relatedness measures by assuming a mutation probability u per individual reproduction (and, at a group reproduction event, all individuals in the group reproduce with the same mutation probability), and then take the limit of $u \downarrow 0$ to find the no-mutation limit of relatedness measures. To do so, we will first derive the recurrence relations for IBD probabilities by assuming a stationary distribution $\{q_i\}$ and then derive the no-mutation limit for relatedness using the following identity (Malécot, 1948; Rousset, 2004; Taylor et al., 2007a;b; Grafen, 2007; Durrett, 2008)

$$r_i = \frac{q_i - \bar{q}}{1 - \bar{q}} \quad (5)$$

where q_i denotes the stationary IBD probability for two individuals whose groups are i steps apart and \bar{q} denotes the average IBD probability of a focal individual to all the individuals in the population, including self.

Two observations will be useful in our later derivations. The first is that, by symmetry, $q_i = q_{m-i}$ for $0 \leq i \leq m$. The second is that, for $i = 0$, we can relate q_0 – the IBD probability for two members from the same group, drawn with replacement – and q_s – the same probability without replacement – in a straightforward way:

$$q_0 = \frac{1 + (n-1)q_s}{n}. \quad (6)$$

since the individual's relatedness to herself is 1.

S.4.1 Birth-Death

In the BD process, an individual can be replaced if:

- An individual event happens, the group the individual is in is chosen to host it, and within the group, she is chosen to be replaced.
- A group event happens, one of the neighbouring groups of the group the individual is in is chosen to reproduce, and that group replaces her group.
- A group event happens, the group that the individual belongs to is chosen, and replaces itself.
- A migration event occurs, the individual's group is chosen to be one of the groups in which the migration takes place, and the individual is chosen to be swapped.

Combining these events, we can derive the recurrence relations for the stationary IBD probabilities.

S.4.1.1 $i = 1$

For $i = 1$, we have the following recurrence relation:

$$\begin{aligned}
 q_1 = & \underbrace{p \frac{2}{m} \frac{1}{n} (1-u) q_1}_{\text{replaced in an individual event}} + \underbrace{q \left(\frac{1}{m} \frac{m-1}{m} (1-u) (q_0 + q_2) + \frac{2}{m^2} (1-u) q_1 \right)}_{\text{replaced in a group event}} + \underbrace{r \frac{1}{m} \frac{1}{n} \left(2q_2 + \frac{2(n-1)}{n} q_s + \frac{1}{n} q_1 \right)}_{\text{replaced in a migration event}} \\
 & + \left(\underbrace{p \left(\frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right)}_{\text{no change through individual events}} + \underbrace{q \left(\frac{m-4}{m} + \frac{m+1}{m^2} + \frac{m-1}{m^2} \right)}_{\text{no change through group events}} + \underbrace{r \left(\frac{m-3}{m} + \frac{2}{m} \frac{n-1}{n} + \frac{1}{m} \frac{(n-1)^2}{n^2} \right)}_{\text{no change through migration events}} \right) q_1
 \end{aligned}$$

Here we use the probabilities with which different replacement events happen. We will sometimes refer to the two individuals in two neighbouring groups as the two focal individuals, and to their groups as the two focal groups.

If an individual event happens, the probability that any given group is chosen to host it is $\frac{1}{m}$. The probability that any individual within that group is chosen to be replaced is $\frac{1}{n}$. Therefore, for two given individuals in neighbouring groups, that adds up to $\frac{2}{m} \frac{1}{n}$. There are a few ways in which they can both not be affected. One of the other $m-2$ groups can be chosen, which happens with probability $\frac{m-2}{m}$; or one of the two neighbouring groups can be chosen, while some other individual is replaced, which happens with probability $\frac{2}{m} \frac{n-1}{n}$.

If a group event happens, then one of the two neighbouring groups can replace the other – which happens with probability $\frac{1}{m} \frac{m-1}{m} \frac{1}{2}$ – or the other can replace the one, which happens with the same probability. In both cases, q_0 is the relevant IBD probability. Also, a neighbour outside the focal pair of groups can replace one of the two in the focal pair, which again happens with a probability that is twice $\frac{1}{m} \frac{m-1}{m} \frac{1}{2}$. In this case, q_2 is the relevant IBD probability. Finally, with a probability of twice $\frac{1}{m^2}$, one of the focal groups replaces itself, in which case, q_1 is the relevant IBD probability. Nothing happens to the pair if a group is chosen to reproduce that is neither of the two groups within the focal pair, nor one of their direct neighbours, which happens with probability $\frac{m-4}{m}$. Also nothing happens if one of the neighbours of the focal pair is chosen, and they replace themselves ($\frac{1}{m^2}$) or their other neighbour ($\frac{1}{m} \frac{m-1}{m} \frac{1}{2}$). These probabilities together times 2 is $\frac{1}{m} \frac{m+1}{m^2}$. Finally, nothing happens if one of the two focal groups is chosen, and replaces a group outside the focal pair, which happens with probability $\frac{1}{m} \frac{m-1}{m^2}$.

With a migration event, every neighbouring pair is doing an exchange with probability $\frac{1}{m}$. This pair consists of both focal groups with probability $\frac{1}{m}$, and it is a pair that consists of one of the two focal groups and its neighbour on the other side with probability $\frac{2}{m}$. In the last case, the focal individual is chosen with probability $\frac{1}{n}$, and q_2 is the relevant IBD probability. In case the exchange is between the focal pair of groups, the two focal individuals themselves are chosen to switch with probability $\frac{1}{n^2}$, one of them is swapped with an individual that is not the other with probability $\frac{n-1}{n^2}$, and the other is swapped with an individual that is not the one with the same probability. In the first case, nothing changes with regard to the IBD probabilities, and in the latter two cases, q_s is the relevant IBD probability. Nothing happens at migration if one of the other $m-3$ pairs is chosen ($\frac{m-3}{m}$), a focal group and a neighbouring group outside the focal pair is chosen, but the focal individual is not chosen ($\frac{2}{m} \frac{n-1}{n}$), or the focal pair itself is chosen, but two individuals other than the focal ones trade places ($\frac{1}{m} \frac{(n-1)^2}{n^2}$).

If we gather the terms with q_1 on the left-hand side, we get

$$\begin{aligned} & \left(1 - p \left(\frac{2}{m} \frac{1}{n} (1-u) + \frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right) - q \left(\frac{2}{m^2} (1-u) + \frac{m-2}{m} \right) \right. \\ & \quad \left. - r \left(\frac{1}{m} \frac{1}{n^2} + \frac{m-3}{m} + \frac{2}{m} \frac{n-1}{n} + \frac{1}{m} \frac{(n-1)^2}{n^2} \right) \right) q_1 \\ & = q \frac{m-1}{m^2} (1-u) (q_0 + q_2) + r \frac{2}{mn} \left(q_2 + \frac{n-1}{n} q_s \right) \end{aligned}$$

If we use the identity $q_0 = \frac{1}{n} + \frac{n-1}{n} q_s$, and multiply left and right by $\frac{m}{2}$, this can be rewritten as

$$\begin{aligned} & \left(\frac{m}{2} - \frac{p}{n} (1-u) - p \left(\frac{m-2}{2} + \frac{n-1}{n} \right) - \frac{q}{m} (1-u) - q \left(\frac{m-2}{2} \right) \right. \\ & \quad \left. - r \left(\frac{1}{2} \frac{1}{n^2} + \frac{m-3}{2} + \frac{n-1}{n} + \frac{1}{2} \frac{(n-1)^2}{n^2} \right) \right) q_1 \\ & = \left(\frac{q}{2} \frac{m-1}{m} (1-u) + \frac{r}{n} \right) (q_0 + q_2) - r \frac{1}{n^2} \end{aligned}$$

If we furthermore use that $p + q + r = 1$, this can be further simplified to

$$\left(1 - (1-u) \left(\frac{p}{n} + \frac{q}{m} \right) - p \frac{n-1}{n} - r \left(\frac{n-1}{n} \right)^2 \right) q_1 = \left((1-u) \frac{q}{2} \frac{m-1}{m} + \frac{r}{n} \right) (q_0 + q_2) - \frac{r}{n^2} \quad (7)$$

S.4.1.2 $1 < i < m-1$

The recurrence relations for $1 < i < m-1$ are derived in a similar way. The differences with the recurrence relation for $i = 1$ arise because the two groups no longer are each other's neighbours, which means that the groups can no longer replace each other, nor can they exchange individuals.

$$\begin{aligned} q_i &= \underbrace{p \frac{2}{m} \frac{1}{n} (1-u) q_i}_{\text{replaced in an individual event}} + \underbrace{q \left(\frac{1}{m} \frac{m-1}{m} (1-u) (q_{i-1} + q_{i+1}) + \frac{2}{m^2} (1-u) q_i \right)}_{\text{replaced in a group event}} + \underbrace{r \frac{2}{m} \frac{1}{n} (q_{i-1} + q_{i+1})}_{\text{replaced in a migration event}} \\ &+ \left(\underbrace{p \left(\frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right)}_{\text{no change through individual events}} + \underbrace{q \left(\frac{m-6}{m} + 2 \frac{m+1}{m^2} + 2 \frac{m-1}{m^2} \right)}_{\text{no change through group events}} + \underbrace{r \left(\frac{m-4}{m} + \frac{4}{m} \frac{n-1}{n} \right)}_{\text{no change through migration events}} \right) q_i \end{aligned}$$

If we gather the terms with q_i on the left-hand side, we get:

$$\begin{aligned} & \left(1 - p \left(\frac{2}{m} \frac{1}{n} (1-u) + \frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right) - q \left(\frac{2}{m^2} (1-u) + \frac{m-2}{m} \right) - r \left(\frac{m-4}{m} + \frac{4}{m} \frac{n-1}{n} \right) \right) q_i \\ & = \left(q \frac{m-1}{m^2} (1-u) + r \frac{2}{mn} \right) (q_{i-1} + q_{i+1}) \end{aligned}$$

If we multiply left and right by $\frac{m}{2}$, and use $p + q + r = 1$, this can be rewritten as

$$\left(\frac{p}{n} + q + \frac{2r}{n} - (1-u) \left(\frac{p}{n} + \frac{q}{m} \right) \right) q_i = \left((1-u) \frac{q}{2} \frac{m-1}{m} + \frac{r}{n} \right) (q_{i-1} + q_{i+1}) \quad (8)$$

S.4.1.3 $i = 0$

To derive q_s – and therewith q_0 – we will use the recurrence relation concerning the IBD probabilities for two individuals within the same group.

$$q_s = p \frac{1}{m} \frac{2}{n^2} (1-u) + p \left(\frac{1}{m} \frac{2(n-1)}{n^2} (1-u) + \frac{1}{m} \frac{n-2}{n} + \frac{m-1}{m} \right) q_s + q \left(\frac{1}{m^2} + \frac{m-1}{m^2} \right) (1-u)^2 q_s \\ + q \left(\frac{m-3}{m} + \frac{m+1}{m^2} + \frac{m-1}{m^2} \right) q_s + r \frac{4}{mn} q_1 + r \left(\frac{m-2}{m} + \frac{2(n-2)}{mn} \right) q_s$$

The relevant probabilities in case of an individual event are $\frac{1}{m} \frac{2}{n^2}$ for the focal group being chosen and one focal individual replacing the other, or vice versa; $\frac{1}{m} \frac{2(n-1)}{n^2}$ for the focal group being chosen, and any individual other than the other focal one replacing one of the focal ones (including herself), or vice versa; $\frac{1}{m} \frac{n-2}{n}$ for the focal group being chosen, and one of the other individuals being replaced; and $\frac{m-1}{m}$ for a different group being chosen.

In case of a group event, the relevant probabilities are $\frac{1}{m^2}$ for the focal group replacing itself; $\frac{m-1}{m^2}$ for the focal group being replaced by a neighbouring group; $\frac{m-3}{m}$ for a group being chosen to reproduce that is not the focal group nor a neighbour; $\frac{m+1}{m^2}$ for a neighbouring group being chosen to reproduce, and replacing itself or it's other neighbour; and $\frac{m-1}{m^2}$ for the focal group being chosen to reproduce, and replacing a neighbouring group and not itself.

In case of a migration event, the relevant probabilities are $\frac{4}{mn}$ for the focal group to exchange a member with the neighbouring group on the right or on the left, and one of the two focal individuals being chosen; $\frac{2(n-2)}{mn}$ for the exchange happening between the focal group and any of the two neighbouring groups, and neither of the two focal individuals being replaced; and $\frac{m-2}{m}$ for the exchange happening in any of the other $m-2$ pairs.

The equation can be rewritten as

$$q_s = p \frac{1}{m} \frac{2}{n^2} (1-u) + p \left(1 - \frac{2}{mn^2} + \frac{2}{mn} \left(1 - \frac{1}{n} \right) (-u) \right) q_s + q \left(\frac{1}{m} \right) (1-u)^2 q_s \\ + q \left(\frac{m-1}{m} \right) q_s + r \frac{4}{mn} q_1 + r \left(1 - \frac{4}{mn} \right) q_s$$

Expressing q_1 as a function of q_s , this becomes

$$q_1 = \frac{mn}{4r} \left(1 - p + p \frac{2}{mn} \left(u + \frac{1-u}{n} \right) - q \left(\frac{m-1}{m} + \frac{(1-u)^2}{m} \right) - r + r \frac{4}{mn} \right) q_s - \frac{p}{r} \frac{1-u}{2n}$$

Finally, we use $p + q + r = 1$ to write

$$q_1 = \frac{mn}{4r} \left(p \frac{2}{mn} \left(u + \frac{1-u}{n} \right) + q \frac{1}{m} - q \frac{(1-u)^2}{m} + r \frac{4}{mn} \right) q_s - \frac{p}{r} \frac{1-u}{2n} \quad (9)$$

S.4.1.4 Solving the system

Suppose, for $1 \leq i \leq m-1$, q_i has the following form:

$$q_i = s_i q_1 \quad (10)$$

where $s_1 = s_{m-1} = 1$, and $\lim_{u \rightarrow 0} s_i = 1$ for $1 < i < m-1$, since all IBD probabilities approach 1 in the limit of no mutation. If we rewrite Equation (8) with the assumption in Equation (10), we get the following equality:

$$\left(\frac{p}{n} + q + \frac{2r}{n} - (1-u)\left(\frac{p}{n} + \frac{q}{m}\right)\right) s_i q_1 = \left((1-u)\frac{q}{2}\frac{m-1}{m} + \frac{r}{n}\right) (s_{i-1} + s_{i+1}) q_1$$

Assuming that $q_1 \neq 0$, this is also:

$$\left(\frac{p}{n} + q + \frac{2r}{n} - (1-u)\left(\frac{p}{n} + \frac{q}{m}\right)\right) s_i = \left((1-u)\frac{q}{2}\frac{m-1}{m} + \frac{r}{n}\right) (s_{i-1} + s_{i+1}) \quad (11)$$

Since $q_i \rightarrow 1$ for all i as $u \downarrow 0$, we cannot directly use Equation (5) to calculate relatedness, as both the numerator and the denominator approach zero. Therefore, we will apply L'Hôpital's rule and calculate relatednesses as:

$$r_i = \frac{q'_i - \bar{q}'}{-\bar{q}'} \quad (12)$$

Here the derivatives are taken with respect to u , and evaluated in the limit of $u \downarrow 0$. In order to determine q'_i and \bar{q}' , we want to find s'_i by taking derivatives on both sides of Equation (11).

$$\begin{aligned} \left(\frac{p}{n} + q + \frac{2r}{n} - (1-u)\left(\frac{p}{n} + \frac{q}{m}\right)\right) s'_i + \left(\frac{p}{n} + \frac{q}{m}\right) s_i &= \left((1-u)\frac{q}{2}\frac{m-1}{m} + \frac{r}{n}\right) (s'_{i-1} + s'_{i+1}) \\ &\quad - \frac{q}{2}\frac{m-1}{m} (s_{i-1} + s_{i+1}) \end{aligned}$$

If we evaluate this in the limit of $u \downarrow 0$, then we can also use that $s_i = 1$ for all i in that limit.

$$\left(\frac{q}{2}\frac{m-1}{m} + \frac{2r}{n}\right) s'_i + \frac{p}{n} + \frac{q}{m} = \left(\frac{q}{2}\frac{m-1}{m} + \frac{r}{n}\right) (s'_{i-1} + s'_{i+1}) - \frac{q}{2}\frac{m-1}{m}$$

This can be reorganized as follows:

$$s'_i - \frac{s'_{i-1} + s'_{i+1}}{2} = -\frac{\frac{p}{n} + q}{q\frac{m-1}{m} + r\frac{2}{n}} \quad (13)$$

Summing both sides of the above equation for $2 \leq i \leq m-2$ yields

$$\begin{aligned} \sum_{i=2}^{m-2} s'_i - \frac{1}{2} \sum_{i=2}^{m-2} (s'_{i-1} + s'_{i+1}) &= -\sum_{i=2}^{m-2} \frac{\frac{p}{n} + q}{q\frac{m-1}{m} + r\frac{2}{n}} \\ \sum_{i=2}^{m-2} s'_i - \frac{1}{2} (s'_1 + s'_{m-1} + s'_2 + s'_{m-2}) - \frac{1}{2} \sum_{i=3}^{m-3} 2s'_i &= -(m-3) \frac{\frac{p}{n} + q}{q\frac{m-1}{m} + r\frac{2}{n}} \\ s'_2 + s'_{m-2} - \frac{1}{2} (s'_2 + s'_{m-2}) &= -(m-3) \frac{\frac{p}{n} + q}{q\frac{m-1}{m} + r\frac{2}{n}} \\ s'_2 &= -(m-3) \frac{\frac{p}{n} + q}{q\frac{m-1}{m} + r\frac{2}{n}} \end{aligned} \quad (14)$$

where we used the identity $s'_1 = s'_{m-1} = 0$ – since s_1 and s_{m-1} are constant – and $s'_2 = s'_{m-2}$ – in the third and the fourth lines, respectively. Using the final equation above and the identity $s'_1 = 0$, we can derive the limit values for all s'_i as given below:

$$s'_i = -(i-1)(m-i-1) \frac{\frac{p}{n} + q}{q^{\frac{m-1}{m}} + r \frac{2}{n}} \quad (15)$$

for $2 \leq i \leq m-2$.

Before we can plug these into the relatedness formula, it will be helpful to write \bar{q}' differently. Since $\bar{q} = \frac{1}{m} \sum_{i=0}^{m-1} q_i$, we can take the derivatives of all the terms separately.

$$\bar{q}' = \frac{1}{m} \left(q'_0 + q'_1 + q'_{m-1} + \sum_{i=2}^{m-2} q_i \right) = \frac{1}{m} \left(q'_0 + q'_1 + q'_{m-1} + \sum_{i=2}^{m-2} (s'_i q_1 + s_i q'_1) \right)$$

Now we can use that $q_1 = 1$ and $s_i = 1$ in the limit of $u \downarrow 0$ for all i , that $q'_1 = q'_{m-1}$, and that $s'_1 = s'_{m-1} = 0$ (because $s_1 = s_{m-1} = 1$ regardless of u) when rewriting \bar{q}' .

$$\begin{aligned} \bar{q}' &= \frac{1}{m} \left(q'_0 + q'_1 + q'_{m-1} + \sum_{i=2}^{m-2} (s'_i + q'_1) \right) = \frac{1}{m} \left(q'_0 + (m-1)q'_1 + \sum_{i=2}^{m-2} s'_i \right) \\ &= \frac{1}{m} \left(q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i \right) \end{aligned}$$

We can now plug this in Equation (5) to get

$$r_0 = \frac{\frac{m-1}{m}(q'_0 - q'_1) - \frac{1}{m} \sum_{i=1}^{m-1} s'_i}{-\frac{1}{m} \left(q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i \right)} = \frac{(m-1)(q'_1 - q'_0) + \sum_{i=1}^{m-1} s'_i}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} \quad (16)$$

$$r_1 = \frac{-\frac{1}{m}(q'_0 - q'_1) - \frac{1}{m} \sum_{i=1}^{m-1} s'_i}{-\frac{1}{m} \left(q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i \right)} = \frac{-(q'_1 - q'_0) + \sum_{i=1}^{m-1} s'_i}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} \quad (17)$$

$$r_i = \frac{s'_i - \frac{1}{m}(q'_0 - q'_1) - \frac{1}{m} \sum_{i=1}^{m-1} s'_i}{-\frac{1}{m} \left(q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i \right)} = r_1 - \frac{ms'_i}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} \quad (18)$$

where

$$\sum_{i=1}^{m-1} s'_i = -\frac{(m-1)(m-2)(m-3)}{6} \frac{\frac{p}{n} + q}{q^{\frac{m-1}{m}} + r \frac{2}{n}} \quad (19)$$

In order to have a formula for relatedness that depends only on the parameters of the model (p , q , r , m and n), we still need to express q'_0 and q'_1 in terms of these parameters. In order to be able to do that, we will first express q'_s in terms of those parameters.

Step 1 is to take the first derivative with respect to u on both sides of Equation (7), and evaluate them at $u = 0$.

$$\left(1 - p - \frac{q}{m} - r \left(\frac{n-1}{n} \right)^2 \right) q'_1 + \left(\frac{p}{n} + \frac{q}{m} \right) q_1 = \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n} \right) (q'_0 + q'_2) - \left(\frac{q}{2} \frac{m-1}{m} \right) (q_0 + q_2)$$

At $u = 0$, $q_0 = q_1 = q_2 = 1$, and hence

$$\left(1 - p - \frac{q}{m} - r \left(\frac{n-1}{n}\right)^2\right) q'_1 + \frac{p}{n} = \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n}\right) (q'_0 + q'_2) - q$$

Also $q'_0 = \frac{n-1}{n} q'_s$ and $q'_2 = s'_2 q_1 + s_2 q'_1$, and hence, with $q_2 = q_1 = 1$ and $s_2 = 1$, at $u = 0$, the expression above becomes

$$\left(1 - p - \frac{q}{m} - r \left(\frac{n-1}{n}\right)^2\right) q'_1 + \frac{p}{n} = \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n}\right) \left(\frac{n-1}{n} q'_s + s'_2 + q'_1\right) - q$$

or

$$\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2 + n}{n^2}\right) q'_1 + \frac{p}{n} = \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n}\right) \left(\frac{n-1}{n} q'_s + s'_2\right) - q$$

Step 2 is to take the first derivative with respect to u in Equation (9)

$$q'_1 = \frac{mn}{4r} \left(p \frac{2}{mn} \left(u + \frac{1-u}{n}\right) + q \frac{1}{m} - q \frac{(1-u)^2}{m} + r \frac{4}{mn}\right) q'_s + \frac{mn}{4r} \left(p \frac{2}{mn} \left(1 - \frac{1}{n}\right) + 2q \frac{(1-u)}{m}\right) q_s + \frac{p}{r} \frac{1}{2n}$$

Evaluated at $u = 0$, where also $q_s = 1$, this is

$$\begin{aligned} q'_1 &= \left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s + \frac{mn}{4r} \left(p \frac{2}{mn} \left(1 - \frac{1}{n}\right) + 2q \frac{1}{m}\right) + \frac{p}{r} \frac{1}{2n} = \left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s + \frac{p}{2r} \left(1 - \frac{1}{n}\right) + \frac{qn}{2r} + \frac{p}{r} \frac{1}{2n} \\ &= \left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s + \frac{p}{2r} + \frac{qn}{2r} \end{aligned}$$

If we combine these two steps, we get

$$\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2 + n}{n^2}\right) \left(\left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s + \frac{p}{2r} + \frac{qn}{2r}\right) + \frac{p}{n} = \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n}\right) \left(\frac{n-1}{n} q'_s + s'_2\right) - q$$

or

$$\begin{aligned} &\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2 + n}{n^2}\right) \left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s - \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n}\right) \frac{n-1}{n} q'_s \\ &= \left(\frac{q}{2} \frac{m-1}{m} + \frac{r}{n}\right) s'_2 - \left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2 + n}{n^2}\right) \left(\frac{p}{2r} + \frac{qn}{2r}\right) - q - \frac{p}{n} \end{aligned}$$

or

$$\begin{aligned} &\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2 + n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) q'_s - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n} q'_s \\ &= \left(q \frac{m-1}{m} + \frac{2r}{n}\right) s'_2 - \left(\frac{n}{r} - \frac{pn}{r} - \frac{qn}{2r} \frac{m+1}{m} - \frac{(n-1)^2 + n}{n}\right) \left(\frac{p}{n} + q\right) - 2 \left(\frac{p}{n} + q\right) \end{aligned}$$

which gives

$$\begin{aligned} q'_s &= \frac{\left(q \frac{m-1}{m} + \frac{2r}{n}\right) s'_2 - \left(\frac{p}{n} + q\right) \left(\frac{n}{r} - \frac{pn}{r} - \frac{qn}{2r} \frac{m+1}{m} - \frac{(n-1)^2}{n} + 1\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{\left(q \frac{m-1}{m} + \frac{2r}{n}\right) s'_2 - \left(\frac{p}{n} + q\right) \left(3 - \frac{1}{n} - \frac{qn}{2r} \frac{m+1}{m} + \frac{n}{r} - \frac{pn}{r} - n\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \end{aligned}$$

We can simplify this using the formula for s'_2 from Equation (14), repeated below.

$$s'_2 = -(m-3) \frac{\frac{p}{n} + q}{q \frac{m-1}{m} + r \frac{2}{n}}$$

If we look at the first term in the numerator in Equation (20), we see that the coefficient of s'_2 is equal to the denominator of s'_2 , so we can rewrite the formula for q'_s as follows:

$$\begin{aligned} q'_s &= \frac{-(m-3) \left(\frac{p}{n} + q\right) - \left(\frac{p}{n} + q\right) \left(3 - \frac{1}{n} - \frac{qn}{2r} \frac{m+1}{m} + \frac{n}{r} - \frac{pn}{r} - n\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} - \frac{qn}{2r} \frac{m+1}{m} + \frac{n}{r} - \frac{pn}{r} - n\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \end{aligned}$$

We can further rewrite the numerator,

$$\begin{aligned} q'_s &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} - \frac{qn}{2r} \frac{m+1}{m} + n \left(\frac{1-p-r}{r}\right)\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} - \frac{qn}{2r} \frac{m+1}{m} + n \frac{q}{r}\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{(n-1)^2+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \end{aligned}$$

where we used $p + q + r = 1$ in the second line.

Then, we can rewrite the denominator,

$$\begin{aligned} q'_s &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r \frac{n^2-2n+1+n}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\left(1 - p - \frac{q}{2} \frac{m+1}{m} - r + r \frac{n-1}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\left(q - \frac{q}{2} \frac{m+1}{m} + r \frac{n-1}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\left(\frac{q}{2} \frac{m-1}{m} + r \frac{n-1}{n^2}\right) \left(\frac{p}{r} \frac{1}{n} + 2\right) - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \\ &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\frac{pq}{2rn} \frac{m-1}{m} + q \frac{m-1}{m} + p \frac{n-1}{n^3} + r \frac{2(n-1)}{n^2} - \left(q \frac{m-1}{m} + \frac{2r}{n}\right) \frac{n-1}{n}} \end{aligned}$$

after which we arrive at

$$q'_s = \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \quad (20)$$

Equation (6) moreover implies

$$q'_0 = \frac{n-1}{n} q'_s \quad (21)$$

which links q'_s to q'_0 , and Step 2 above gave us

$$q'_1 = \left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s + \frac{p+nq}{2r} \quad (22)$$

which links q'_s to q'_1 . This implies that we have everything we need to complete Equations (16), (17) and (18) for the BD process.

S.4.2 Shift

In the Shift process, an individual can be replaced if:

- An individual event happens, the group the individual is in is chosen to host it, and then within the group, she is chosen to be replaced.
- A group event happens, and a neighbouring group, or its offspring group, pushes the individual's group one position away from where it was, or replaces it. Probabilities for those events are also derived in Allen and Nowak (2012) for the Shift process, where all positions are occupied by individuals instead of groups.
- A migration event occurs, the individual's group is chosen to be one of the groups in which the migration takes place, and the individual is chosen to be swapped.

Combining these events, we can derive the recurrence relations for the stationary IBD probabilities.

S.4.2.1 $i = 1$

For $i = 1$, we have the following recurrence relation:

$$\begin{aligned} q_1 = & \underbrace{p \frac{2}{m} \frac{1}{n} (1-u) q_1}_{\text{replaced in an individual event}} + \underbrace{q \left(\frac{(m-1)(1-u)}{m^2} q_0 + \frac{1-u}{m} q_1 + \frac{(m-2+(1-u))}{m^2} q_2 \right)}_{\text{replaced in a group event}} \\ & + \underbrace{r \frac{1}{m} \frac{1}{n} \left(2q_2 + \frac{2(n-1)}{n} q_s + \frac{1}{n} q_1 \right)}_{\text{replaced in a migration event}} \\ & + \left(\underbrace{p \left(\frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right)}_{\text{no change through individual events}} + \underbrace{q \left(\frac{(m-1)(m-2)}{m^2} \right)}_{\text{no change through group events}} + \underbrace{r \left(\frac{m-3}{m} + \frac{2}{m} \frac{n-1}{n} + \frac{1}{m} \frac{(n-1)^2}{n^2} \right)}_{\text{no change through migration events}} \right) q_1 \end{aligned}$$

The probabilities for individual and migration events are the same as in BD. The probabilities for group events are different. To get the probabilities for group events right, it is important, in the face of equivalent ways to define this update rule, to have an unambiguous rule for who ends up at which location after a group reproduction event. The reproducing group always stays put. If it is also chosen to die, its offspring group takes its place and no group moves. If not, then with probability one half, the offspring group occupies the position to the left of the parent group, and every group in between the reproducing and the dying group moves in the same direction. Also with probability one half, the offspring group occupies the position on the right, and every group in between the reproducing and the dying group moves in that direction.

Now, for a given pair of neighbouring positions, the group on the left replaces the group on the right, if the left one is chosen to reproduce, not chosen to also die, and reproduces on the right. The right one replaces the left one if the right one is chosen to reproduce, not chosen to also die, and reproduces to the left. Both events happen with probability $\frac{1}{2} \frac{m-1}{m^2}$, and after this, two randomly chosen members of the neighbouring groups are IBD with probability $(1-u)q_0$.

The left one replaces itself with probability $\frac{1}{m^2}$. The right one too. After this, two randomly chosen members of the neighbouring groups are IBD with probability $(1-u)q_1$.

The group to the left of the left one reproduces to the right and pushes the left group to the right position if the group to the left of the left one is chosen to reproduce, the left group and the group to the left of the left one are both not chosen to die, and the reproducing group reproduces to the right. This happens with probability $\frac{1}{2} \frac{m-2}{m^2}$. The mirror image of that happens with the same probability. After this, two randomly chosen members of the neighbouring groups are IBD with probability $(1-u)q_1$.

These probabilities and the probabilities with which they replace themselves add up to $\frac{1}{m}$, and we have seen that, for both, the IBD probability is $(1-u)q_1$.

The left one is replaced by the group to the left of it, if the left one is chosen to die, any group other than the left one and its left neighbour is chosen to reproduce, and the reproducing group reproduces to the right. The right one is replaced by the group to the right of it, if the right one is chosen to die, any group other than the right one and its right neighbour is chosen to reproduce, and the reproducing group reproduces to the left. Both events happen with probability $\frac{1}{2} \frac{m-2}{m^2}$, and after this, two randomly chosen members of the neighbouring groups are IBD with probability q_2 .

The left one is replaced by the offspring of the group to the left of it, if the left one is chosen to die, its left neighbour is chosen to reproduce, and it reproduces to the right. The right one is replaced by the offspring of the group to the right of it, if the right one is chosen to die, its right neighbour is chosen to reproduce, and it reproduces to the left. Both events happen with probability $\frac{1}{2} \frac{1}{m^2}$, and after this, two randomly chosen members of the neighbouring groups are IBD with probability $(1-u)q_2$.

After all other events, the groups at the two given neighbouring locations are both not the offspring group, and were neighbouring groups in the period before the group event, too.

All of those group event probabilities can also be found in Allen and Nowak (2012). The only differences are that we derive them in a forward looking way, while they do it in a backward looking way, and, since we may have more than one individual at any site, we do not have $q_0 = 1$.

If we gather all terms with q_1 on the left hand side, we get

$$\begin{aligned} & \left(1 - p \left(\frac{2}{m} \frac{1}{n} (1-u) + \frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right) - q \left(\frac{1-u}{m} + \frac{(m-1)(m-2)}{m^2} \right) \right. \\ & \quad \left. - r \left(\frac{1}{m} \frac{1}{n^2} + \frac{m-3}{m} + \frac{2}{m} \frac{n-1}{n} + \frac{1}{m} \frac{(n-1)^2}{n^2} \right) \right) q_1 = \\ & q \left(\frac{(m-1)(1-u)}{m^2} q_0 + \frac{(m-2+(1-u))}{m^2} q_2 \right) + r \frac{1}{m} \frac{1}{n} \left(2q_2 + \frac{2(n-1)}{n} q_s \right) \end{aligned}$$

If we use the identity $q_0 = \frac{1+(n-1)q_s}{n}$, and multiply left and right by m , this can be rewritten as

$$\begin{aligned} & \left(m - p \left(m - \frac{2}{n} u \right) - q \left(m - \frac{2m-2}{m} - u \right) - r \left(m - \frac{4n-2}{n^2} \right) \right) q_1 = \\ & \left(q \frac{m-1}{m} + r \frac{2}{n} \right) (q_0 + q_2) - q \frac{u}{m} ((m-1)q_0 + q_2) - r \frac{2}{n^2} \end{aligned}$$

If we furthermore use that $p + q + r = 1$, this can be further simplified to

$$\left(q \frac{2(m-1)}{m} + r \frac{2(2n-1)}{n^2} + u \left(p \frac{2}{n} + q \right) \right) q_1 = \left(q \frac{m-1}{m} + r \frac{2}{n} \right) (q_0 + q_2) - q \frac{u}{m} ((m-1)q_0 + q_2) - r \frac{2}{n^2} \quad (23)$$

S.4.2.2 $1 < i < m-1$

The recurrence relations for $1 < i < m-1$ are derived in a similar way. The differences with the recurrence relation for $i = 1$ arise because the two groups no longer are each other's neighbours, which means that the groups can no longer replace each other, nor can they swap individuals.

$$\begin{aligned} q_i = & \underbrace{p \frac{2}{m} \frac{1}{n} (1-u) q_i}_{\text{replaced in an individual event}} + \underbrace{q \left(\frac{(m-i)(i-1+(1-u))}{m^2} q_{i-1} + \frac{1-u}{m} q_i + \frac{i(m-i-1+(1-u))}{m^2} q_{i+1} \right)}_{\text{replaced in a group event}} \\ & + \underbrace{r \frac{2}{m} \frac{1}{n} (q_{i-1} + q_{i+1})}_{\text{replaced in a migration event}} \\ & + \left(\underbrace{p \left(\frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right)}_{\text{no change through individual events}} + \underbrace{q \left(\frac{(m-i)(m-i-1) + i(i-1)}{m^2} \right)}_{\text{no change through group events}} + \underbrace{r \left(\frac{m-4}{m} + \frac{4}{m} \frac{n-1}{n} \right)}_{\text{no change through migration events}} \right) q_i \end{aligned}$$

If we gather the terms with q_i on the left-hand side, we get

$$\begin{aligned} & \left(1 - p \left(\frac{2}{m} \frac{1}{n} (1-u) + \frac{m-2}{m} + \frac{2}{m} \frac{n-1}{n} \right) - q \frac{(m-i)(m-i-u) + i(i-u)}{m^2} - r \left(\frac{m-4}{m} + \frac{4}{m} \frac{n-1}{n} \right) \right) q_i \\ & = \left(q \frac{(m-i)(i-u)}{m^2} + r \frac{2}{m} \frac{1}{n} \right) q_{i-1} + \left(q \frac{i(m-i-u)}{m^2} + r \frac{2}{m} \frac{1}{n} \right) q_{i+1} \end{aligned}$$

If we multiply both sides with m , and use $p + q + r = 1$ again, we can rewrite this as

$$\left(q \frac{2i(m-i)}{m} + r \frac{4}{n} + u \left(p \frac{2}{n} + q \right) \right) q_i = \left(q \frac{i(m-i)}{m} + r \frac{2}{n} \right) (q_{i-1} + q_{i+1}) - q \frac{u}{m} ((m-i)q_{i-1} + i q_{i+1}) \quad (24)$$

S.4.2.3 $i = 0$

To derive q_s – and therewith q_0 – we use the recurrence relation concerning the IBD probabilities for two individuals from one and the same group.

$$\begin{aligned} q_s = & p \frac{1}{m} \frac{2}{n^2} (1-u) + p \left(\frac{1}{m} \frac{2(n-1)}{n^2} (1-u) + \frac{1}{m} \frac{n-2}{n} + \frac{m-1}{m} \right) q_s + q \left(\frac{(1-u)^2}{m} + \frac{m-1}{m} \right) q_s \\ & + r \frac{4}{mn} q_1 + r \left(\frac{m-2}{m} + \frac{2(n-2)}{mn} \right) q_s \end{aligned}$$

Again, the probabilities for individual and migration events are the same as in BD. If a group event happens, then any given group reproduces and replaces itself with probability $\frac{1}{m^2}$, is replaced by its left neighbour with probability $\frac{1}{2} \frac{m-1}{m^2}$, and by its right neighbour with the same probability. That adds up to $\frac{1}{m}$, and the IBD probability dilutes to $(1-u)^2 q_s$ in all of these cases. In all other cases, it remains q_s .

Expressing q_1 as a function of q_s , this becomes

$$q_1 = \frac{mn}{4r} \left(1 - p + p \frac{2}{mn} \left(u + \frac{1-u}{n} \right) - q \left(\frac{(1-u)^2}{m} + \frac{m-1}{m} \right) - r + r \frac{4}{mn} \right) q_s - \frac{p}{r} \frac{1-u}{2n}$$

Finally, we use $p + q + r = 1$ to write

$$q_1 = \frac{mn}{4r} \left(p \frac{2}{mn} \left(u + \frac{1-u}{n} \right) + q \frac{1}{m} - q \frac{(1-u)^2}{m} + r \frac{4}{mn} \right) q_s - \frac{p}{r} \frac{1-u}{2n} \quad (25)$$

This turns out to be the same equation for Shift as for BD.

S.4.2.4 Solving the system

Suppose again that q_i has the following form for $1 \leq i \leq m-1$:

$$q_i = s_i q_1$$

where $s_1 = s_{m-1} = 1$, and $\lim_{u \rightarrow 0} s_i = 1$ for $1 < i < m-1$. Then, we can rewrite Equation (24) as follows:

$$\left(q \frac{2i(m-i)}{m} + r \frac{4}{n} + u \left(p \frac{2}{n} + q \right) \right) s_i q_1 = \left(q \frac{i(m-i)}{m} + r \frac{2}{n} \right) (s_{i-1} + s_{i+1}) q_1 - q \frac{u}{m} ((m-i)s_{i-1} + i s_{i+1}) q_1$$

Assuming that $q_1 \neq 0$, this is also:

$$\left(q \frac{2i(m-i)}{m} + r \frac{4}{n} + u \left(p \frac{2}{n} + q \right) \right) s_i = \left(q \frac{i(m-i)}{m} + r \frac{2}{n} \right) (s_{i-1} + s_{i+1}) - q \frac{u}{m} ((m-i)s_{i-1} + i s_{i+1}) \quad (26)$$

Since $q_i \rightarrow 1$ for all i as $u \downarrow 0$, we cannot directly use Equation (5) to calculate relatedness, as both the numerator and the denominator approach zero. Therefore, we will apply L'Hôpital's rule and calculate

relatednesses as:

$$r_i = \frac{q'_i - \bar{q}'}{-\bar{q}'}$$

Here the derivatives are taken with respect to u , and evaluated in the limit of $u \downarrow 0$. In order to determine q'_i and \bar{q}' , we want to find s'_i by taking derivatives with respect to u on both sides of Equation (26).

$$\begin{aligned} & \left(q \frac{2i(m-i)}{m} + r \frac{4}{n} + u \left(p \frac{2}{n} + q \right) \right) s'_i + \left(p \frac{2}{n} + q \right) s_i = \\ & \left(q \frac{i(m-i)}{m} + r \frac{2}{n} \right) (s'_{i-1} + s'_{i+1}) - \frac{q}{m} ((m-i)s_{i-1} + is_{i+1}) - q \frac{u}{m} ((m-i)s'_{i-1} + is'_{i+1}) \end{aligned}$$

If we evaluate this in the limit of $u \downarrow 0$, then we can also use that $s_i = 1$ for all i in that limit.

$$\left(q \frac{2i(m-i)}{m} + r \frac{4}{n} \right) s'_i + \left(p \frac{2}{n} + q \right) = \left(q \frac{i(m-i)}{m} + r \frac{2}{n} \right) (s'_{i-1} + s'_{i+1}) - q$$

This can be reorganized as follows

$$\left(q \frac{i(m-i)}{m} + r \frac{2}{n} \right) (2s'_i - (s'_{i-1} + s'_{i+1})) = - \left(p \frac{2}{n} + 2q \right)$$

If we divide everything by 2 and by the first term, we get the following equation.

$$s'_i - \frac{s'_{i-1} + s'_{i+1}}{2} = - \frac{\frac{p}{n} + q}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \quad (27)$$

We will call the right hand side of this equation $-\eta_i$. If we do so, and we sum both sides of the equation over $2 \leq i \leq m-2$, we get

$$\begin{aligned} & \sum_{i=2}^{m-2} s'_i - \frac{1}{2} \sum_{i=2}^{m-2} (s'_{i-1} + s'_{i+1}) = - \sum_{i=2}^{m-2} \eta_i \\ & \sum_{i=2}^{m-2} s'_i - \frac{1}{2} (s'_1 + s'_{m-1} + s'_2 + s'_{m-2}) - \frac{1}{2} \sum_{i=3}^{m-3} 2s'_i = - \sum_{i=2}^{m-2} \eta_i \\ & s'_2 + s'_{m-2} - \frac{1}{2} (s'_2 + s'_{m-2}) = - \sum_{i=2}^{m-2} \eta_i \end{aligned}$$

$$s'_2 = - \sum_{i=2}^{m-2} \eta_i \quad (28)$$

which, together with Equation (27), implies that

$$s'_j = -(j-1) \sum_{i=2}^{m-2} \eta_i + 2 \sum_{i=2}^{j-1} (j-i) \eta_i$$

The derivations of Equations (16), (17) and (18) in the previous subsection do not depend on the update process; they only depend on the assumption that $q_i = s_i q_1$ for $1 \leq i \leq m-1$. Here we make the same

assumption for Shift. Hence, we can use the same equations for Shift as we did for BD. Because Equations (16), (17) and (18) also feature the sum of s'_j values, it will help to compute that sum as well. Since $s'_1 = s'_{m-1} = 0$, summing from $j = 2$ to $m - 2$ gives the same answer as summing from $j = 1$ to $m - 1$.

$$\sum_{j=1}^{m-1} s'_j = \sum_{j=2}^{m-2} s'_j = - \sum_{j=2}^{m-2} (j-1) \sum_{i=2}^{m-2} \eta_i + 2 \sum_{j=2}^{m-2} \sum_{i=2}^{j-1} (j-i) \eta_i$$

The first part of the above summation can be calculated as follows

$$\sum_{j=2}^{m-2} (j-1) \sum_{i=2}^{m-2} \eta_i = (1 + 2 + \dots + m-3) \sum_{i=2}^{m-2} \eta_i = \frac{(m-3)(m-2)}{2} \sum_{i=2}^{m-2} \eta_i \quad (29)$$

The second part of the summation can be calculated as follows

$$\sum_{j=2}^{m-2} \sum_{i=2}^{j-1} (j-i) \eta_i = \underbrace{\eta_2}_{j=3} + \underbrace{2\eta_2 + \eta_3}_{j=4} + \underbrace{3\eta_2 + 2\eta_3 + \eta_4}_{j=5} + \dots + \underbrace{(m-4)\eta_2 + (m-5)\eta_3 + \dots + 3\eta_{m-5} + 2\eta_{m-4} + \eta_{m-3}}_{j=m-2}$$

When we extend the summation in this way, we see that for every η_i , the first time it appears, it is multiplied by 1; the next time it is multiplied by 2; and so on; and the last point the same term appears, it is multiplied by $(m-2-i)$. Therefore, we can rewrite the above summation as follows

$$\sum_{j=2}^{m-2} \sum_{i=2}^{j-1} (j-i) \eta_i = \sum_{i=2}^{m-3} (1 + 2 + \dots + m-2-i) \eta_i = \sum_{i=2}^{m-3} \frac{(m-2-i)(m-1-i)}{2} \eta_i \quad (30)$$

Now, if we combine the two parts found in (29) and (30), we see that

$$\begin{aligned} \sum_{j=2}^{m-2} s'_j &= - \sum_{j=2}^{m-2} (j-1) \sum_{i=2}^{m-2} \eta_i + 2 \sum_{j=2}^{m-2} \sum_{i=2}^{j-1} (j-i) \eta_i \\ &= - \frac{(m-3)(m-2)}{2} \sum_{i=2}^{m-2} \eta_i + 2 \sum_{i=2}^{m-3} \frac{(m-2-i)(m-1-i)}{2} \eta_i \end{aligned}$$

We can write the terms with η_2 and η_{m-2} separately, and use that $\eta_2 = \eta_{m-2}$:

$$\begin{aligned} \sum_{j=2}^{m-2} s'_j &= - \frac{(m-3)(m-2)}{2} (\eta_2 + \eta_{m-2}) - \frac{(m-3)(m-2)}{2} \sum_{i=3}^{m-3} \eta_i + (m-4)(m-3)\eta_2 \\ &\quad + 2 \sum_{i=3}^{m-3} \frac{(m-2-i)(m-1-i)}{2} \eta_i \\ &= -(m-3)(m-2)\eta_2 - \frac{(m-3)(m-2)}{2} \sum_{i=3}^{m-3} \eta_i + (m-4)(m-3)\eta_2 + 2 \sum_{i=3}^{m-3} \frac{(m-2-i)(m-1-i)}{2} \eta_i \\ &= -2(m-3)\eta_2 + \sum_{i=3}^{m-3} \left(2 \frac{(m-2-i)(m-1-i)}{2} - \frac{(m-3)(m-2)}{2} \right) \eta_i \end{aligned}$$

Rearranging gives the formula below.

$$\sum_{j=1}^{m-1} s'_j = \sum_{j=2}^{m-2} s'_j = -2(m-3)\eta_2 + \sum_{i=3}^{m-3} \left(\frac{m^2 - m - 2 - 2i(2m-3) + 2i^2}{2} \right) \eta_i \quad (31)$$

Unfortunately, there is no closed-form solution for the values s'_i and their sum. However, it is possible to find numerical solutions once we fix the population size.

In order to have a formula for relatedness that only depends on the parameters of the model (p, q, r, m and n), we still need to express q'_0 and q'_1 in terms of these parameters. In order to be able to do that, we will first express q'_s in terms of those parameters.

Step 1 is to take the first derivative with respect to u on both sides of Equation (23), and evaluate them at $u = 0$.

$$\left(q \frac{2(m-1)}{m} + r \frac{2(2n-1)}{n^2} \right) q'_1 + \left(p \frac{2}{n} + q \right) q_1 = \left(q \frac{m-1}{m} + r \frac{2}{n} \right) (q'_0 + q'_2) - \frac{q}{m} ((m-1)q_0 + q_2)$$

At $u = 0$, also $q_0 = q_1 = q_2 = 1$, and hence

$$\left(q \frac{2(m-1)}{m} + r \frac{2(2n-1)}{n^2} \right) q'_1 + \left(p \frac{2}{n} + 2q \right) = \left(q \frac{m-1}{m} + r \frac{2}{n} \right) (q'_0 + q'_2)$$

Also $q'_0 = \frac{n-1}{n} q'_s$ and $q'_2 = s'_2 q_1 + s_2 q'_1$, and hence, with $q_2 = q_1 = 1$ and $s_2 = 1$ at $u = 0$, this is also

$$\left(q \frac{2(m-1)}{m} + r \frac{2(2n-1)}{n^2} \right) q'_1 + \left(p \frac{2}{n} + 2q \right) = \left(q \frac{m-1}{m} + r \frac{2}{n} \right) \left(\frac{n-1}{n} q'_s + s'_2 + q'_1 \right)$$

or

$$\left(q \frac{m-1}{m} + r \frac{2(n-1)}{n^2} \right) q'_1 + \left(p \frac{2}{n} + 2q \right) = \left(q \frac{m-1}{m} + r \frac{2}{n} \right) \left(\frac{n-1}{n} q'_s + s'_2 \right)$$

Step 2 is to take the first derivative with respect to u in Equation (25)

$$q'_1 = \frac{mn}{4r} \left(p \frac{2}{mn} \left(u + \frac{1-u}{n} \right) + q \frac{1}{m} - q \frac{(1-u)^2}{m} + r \frac{4}{mn} \right) q'_s + \frac{mn}{4r} \left(p \frac{2}{mn} \left(1 - \frac{1}{n} \right) + 2q \frac{(1-u)}{m} \right) q_s + \frac{p}{r} \frac{1}{2n}$$

Evaluated at $u = 0$, where also $q_s = 1$, this is

$$\begin{aligned} q'_1 &= \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{mn}{4r} \left(p \frac{2}{mn} \left(1 - \frac{1}{n} \right) + 2q \frac{1}{m} \right) + \frac{p}{r} \frac{1}{2n} = \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p}{2r} \left(1 - \frac{1}{n} \right) + \frac{qn}{2r} + \frac{p}{r} \frac{1}{2n} \\ &= \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p}{2r} + \frac{qn}{2r} \end{aligned}$$

If we combine these two steps, we get

$$\left(q \frac{m-1}{m} + r \frac{2(n-1)}{n^2} \right) \left(\left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p}{2r} + \frac{qn}{2r} \right) + \left(p \frac{2}{n} + 2q \right) = \left(q \frac{m-1}{m} + r \frac{2}{n} \right) \left(\frac{n-1}{n} q'_s + s'_2 \right)$$

or

$$\begin{aligned} & \left(q \frac{m-1}{m} + r \frac{2(n-1)}{n^2} \right) \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s - \left(q \frac{m-1}{m} + r \frac{2}{n} \right) \frac{n-1}{n} q'_s = \\ & \left(q \frac{m-1}{m} + r \frac{2}{n} \right) s'_2 - \left(q \frac{m-1}{m} + r \frac{2(n-1)}{n^2} \right) \left(\frac{p}{2r} + \frac{qn}{2r} \right) - \left(\frac{p}{n} + 2q \right) \end{aligned}$$

or

$$\begin{aligned} & \left(\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{m-1}{m} \frac{1}{n} + p \frac{n-1}{n^3} \right) q'_s = \\ & \left(q \frac{m-1}{m} + r \frac{2}{n} \right) s'_2 - \left(\frac{q}{r} \frac{n}{2} \frac{m-1}{m} + \frac{(n-1)}{n} \right) \left(\frac{p}{n} + q \right) - \left(\frac{p}{n} + 2q \right) \end{aligned}$$

which gives

$$q'_s = \frac{\left(q \frac{m-1}{m} + r \frac{2}{n} \right) s'_2 - \left(\frac{p}{n} + q \right) \left(3 - \frac{1}{n} + \frac{q}{r} \frac{n}{2} \frac{m-1}{m} \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}}$$

If we plug in the formula for s'_2 from (28), we get the following expression for q'_s :

$$\begin{aligned} q'_s &= \frac{- \left(q \frac{m-1}{m} + r \frac{2}{n} \right) \sum_{i=2}^{m-2} \frac{\frac{p}{n} + q}{q \frac{i(m-i)}{m} + r \frac{2}{n}} - \left(\frac{p}{n} + q \right) \left(3 - \frac{1}{n} + \frac{q}{r} \frac{n}{2} \frac{m-1}{m} \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\ &= \frac{- \left(\frac{p}{n} + q \right) \left(\sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} + 3 - \frac{1}{n} + \frac{q}{r} \frac{n}{2} \frac{m-1}{m} \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \end{aligned} \quad (32)$$

As in BD, Equation (6) moreover implies

$$q'_0 = \frac{n-1}{n} q'_s \quad (33)$$

which links q'_s to q'_0 , and Step 2 above gave us

$$q'_1 = \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p+nq}{2r} \quad (34)$$

which links q'_s to q'_1 . This implies that we have everything we need to complete Equations (16), (17) and (18) for the Shift process.

S.4.3 Three useful identities

The definition of relatedness (Equation 5) implies that relatednesses have to add up to 0.

$$\sum_{i=0}^{m-1} r_i = \sum_{i=0}^{m-1} \frac{q_i - \bar{q}}{1 - \bar{q}} = \frac{m\bar{q} - m\bar{q}}{1 - \bar{q}} = 0$$

Equation (6) relates the relatedness within the group including self and excluding self in an obvious way, which we repeat here:

$$r_0 = \frac{1 + (n-1)r_s}{n}$$

Together, these imply that

$$\sum_{i=1}^{m-1} r_i = -r_0 = -\frac{1}{n} - \frac{(n-1)r_s}{n} \quad (35)$$

If we define r_o as the relatedness found through IBD probabilities, as in the previous subsection, to a randomly drawn individual from another group, where all other groups are equally likely to be drawn, then this is

$$r_o = \frac{1}{m-1}r_1 + \frac{1}{m-1}r_2 + \frac{1}{m-1}r_3 + \cdots + \frac{1}{m-1}r_{m-1} = \frac{1}{m-1} \sum_{i=1}^{m-1} r_i$$

or

$$\sum_{i=1}^{m-1} r_i = (m-1)r_o \quad (36)$$

Combining Equations (35) and (36), we get

$$r_o = -\frac{1}{n(m-1)} - \frac{(n-1)r_s}{n(m-1)} \quad (37)$$

None of the three identities depends on the update process, so they apply to BD as well as Shift.

S.4.4 Alternative derivation of the three identities

In the limit of $u \downarrow 0$, if two individuals are identical, they are identical by descent. Therefore, if we derive these identities more generally, using conditional probabilities, then they will coincide in this limit. Consider the dynamical system at hand, which is a Markov chain, in which every state is a vector k , where $k_i \in \{0, 1, \dots, n\}$ is the number of cooperators in group i , for $i = 1, \dots, m$. For every such population state, one can imagine hypothetical chance experiments, and define differences in conditional probabilities as we will below. These can then be aggregated, with weights attached to the population states. The weights could represent how often these states are visited relative to each other (this would be the rare-mutation dimorphic distribution from Allen and Tarnita, 2014, or the rare-mutation conditional distribution from Allen and McAvoy, 2018), but for now, all that matters is that p_k is the weight of population state k , and that $\sum_k p_k = 1$.

S.4.4.1 Within-group relatedness

Consider the following hypothetical chance experiment for a given population state k . Draw a random individual from the population, with every individual equally likely to be drawn. After this, go back to the same group, and randomly draw another individual from it. Then, for this state, one could define the proto-relatedness as

$$r_{s,k} = \mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D)$$

The subscript s for same group indicates that this measure is about the relatedness between two different individuals within the same group. The subscript k indicates which population state it pertains to.

Within group relatedness can now be defined as

$$r_s = \sum_k p_k \cdot r_{s,k} = \sum_k p_k (\mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D))$$

S.4.4.2 Relatedness with an individual from a random other group

Now think of another chance experiment for a given population state k . Draw a random individual from the population, again with every individual equally likely to be drawn. After this, go to a different group, with all other groups equally likely to be chosen, and randomly draw another individual from that group. Then, for this state, proto-relatedness is

$$r_{o,k} = \mathbb{P}_{o,k}(C|C) - \mathbb{P}_{o,k}(C|D)$$

The subscript o for other group indicates that this measure is about the relatedness between two individuals in different groups.

Relatedness between two individuals from randomly chosen different groups can now be defined as

$$r_o = \sum_k p_k \cdot r_{o,k} = \sum_k p_k (\mathbb{P}_{o,k}(C|C) - \mathbb{P}_{o,k}(C|D))$$

S.4.4.3 Relatedness between individuals that are i groups apart

With groups situated on the cycle, we can also define other chance experiments for a given population state k . First draw a random individual from the population, as before. After this, with probability $\frac{1}{2}$ go to the group that is i steps to the left of the first group, and with probability $\frac{1}{2}$ go to the group that is i steps to the right of the first group, $i = 1, \dots, m-1$. Randomly draw another individual from that group. Then, for this state, the i -step proto-relatedness, can now be defined as

$$r_{i,k} = \mathbb{P}_{i,k}(C|C) - \mathbb{P}_{i,k}(C|D)$$

for $i = 1, \dots, m-1$, where $r_i = r_j$ if $i = m-j$ for all $1 \leq i \leq m-1$. The subscript i indicates that this measure is about the relatedness between two individuals in groups that are i steps away from each other.

Relatedness between two individuals that are i groups apart can now be defined as

$$r_i = \sum_k p_k \cdot r_{i,k} = \sum_k p_k (\mathbb{P}_{i,k}(C|C) - \mathbb{P}_{i,k}(C|D))$$

S.4.4.4 Identity I

In order to relate these relatednesses to each other, assume that we consider a state k , for which $K = \sum_{j=0}^m k_j$ is the total number of cooperators in the population as a whole. Now imagine a chance experiment where we first draw a random individual from the population, with all individuals equally likely to be chosen, and then – without replacement – another individual from the population as a whole, with all remaining individuals

equally likely to be drawn. Conditional on the first being a cooperator, the chance that the second is a cooperator can be written in two different ways, which must be equal to each other:

$$\frac{n-1}{nm-1}\mathbb{P}_{s,k}(C|C) + \frac{n(m-1)}{nm-1}\mathbb{P}_{o,k}(C|C) = \frac{K-1}{nm-1}$$

Conditional on the first being a defector, one can also express the chance that the second is a cooperator in two equivalent ways

$$\frac{n-1}{nm-1}\mathbb{P}_{s,k}(C|D) + \frac{n(m-1)}{nm-1}\mathbb{P}_{o,k}(C|D) = \frac{K}{nm-1}$$

These two identities together imply that

$$(n-1)(\mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D)) + n(m-1)(\mathbb{P}_{o,k}(C|C) - \mathbb{P}_{o,k}(C|D)) = -1$$

which implies that

$$r_{o,k} = -\frac{1}{n(m-1)} - \frac{n-1}{n(m-1)}r_{s,k}$$

Because the p_k add up to 1, this state-wise identity implies that if we aggregate over states accordingly, the following holds:

$$r_o = -\frac{1}{n(m-1)} - \frac{n-1}{n(m-1)}r_s$$

This is Equation (37).

S.4.4.5 Identity II

One can also go over the groups, according to their distance to the group from which the first individual was drawn. Then the equalities become:

$$\frac{n-1}{nm-1}\mathbb{P}_{s,k}(C|C) + \frac{n}{nm-1} \sum_{i=1}^{m-1} \mathbb{P}_{i,k}(C|C) = \frac{K-1}{nm-1}$$

and

$$\frac{n-1}{nm-1}\mathbb{P}_{s,k}(C|D) + \frac{n}{nm-1} \sum_{i=1}^{m-1} \mathbb{P}_{i,k}(C|D) = \frac{K}{nm-1}$$

These two identities together imply that

$$(n-1)(\mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D)) + n \left(\sum_{i=1}^{m-1} \mathbb{P}_{i,k}(C|C) - \sum_{i=1}^{m-1} \mathbb{P}_{i,k}(C|D) \right) = -1$$

which can be rewritten as

$$\sum_{i=1}^{m-1} r_{i,k} = -\frac{1}{n} - \frac{n-1}{n}r_{s,k}$$

Because the p_k add up to 1, this state-wise identity implies that if we aggregate over states accordingly, the following holds:

$$\sum_{i=1}^{m-1} r_i = -\frac{1}{n} - \frac{n-1}{n} r_s$$

This is Equation (35).

S.4.4.6 Identity III

Given Identity I, Identity II is also equivalent to

$$\sum_{i=1}^{m-1} r_i = (m-1)r_o$$

This is Equation (36).

S.5 Birth-Death versus Shift

In Section S.3, we have found the critical b/c ratios for the the Birth-Death and the Shift process. For the BD process, the critical ratio was given in Equation (3). It is repeated below, and we use the identity $r_0 = \frac{1+(n-1)r_s}{n}$ to rewrite the right hand side.

$$\begin{aligned}\frac{b}{c} &> \frac{p}{q} \frac{n-1}{n} \frac{m}{m-1} \frac{1-r_s}{1+(n-1)r_s - nr_1} \\ &= \frac{p}{q} \frac{1}{n} \frac{m}{m-1} \frac{1-r_0}{r_0 - r_1}\end{aligned}$$

For the Shift process, the critical ratio was given in Equation (4). This is also repeated below, and again we use the identity $r_0 = \frac{1+(n-1)r_s}{n}$ to rewrite the right hand side.

$$\begin{aligned}\frac{b}{c} &> \frac{p}{q} \frac{n-1}{n} \frac{1-r_s}{1+(n-1)r_s} \\ &= \frac{p}{q} \frac{1}{n} \frac{1-r_0}{r_0}\end{aligned}$$

These thresholds are expressed as functions of p , q , r , m , and n , as well as relatednesses r_0 and r_1 – or, equivalently, r_s and r_1 . These relatednesses will typically be different for different update processes.

In Section S.4, we have computed r_0 and r_1 , both for the Birth-Death process and for the Shift process, expressing them as functions of p , q , r , m and n as well. For both processes, Equations (16) and (17), reproduced below, apply.

$$\begin{aligned}r_0 &= \frac{(m-1)(q'_1 - q'_0) + \sum_{i=1}^{m-1} s'_i}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} \\ r_1 &= \frac{-(q'_1 - q'_0) + \sum_{i=1}^{m-1} s'_i}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i}\end{aligned}$$

In these formulas, we still need to fill in q'_1 , q'_0 and $\sum_{i=1}^{m-1} s'_i$, and these will differ between the two processes. For BD, we have Equations (19), (20), (21) and (22), reproduced below.

$$\begin{aligned}\sum_{i=1}^{m-1} s'_i &= -\frac{(m-1)(m-2)(m-3)}{6} \frac{\frac{p}{n} + q}{q^{\frac{m-1}{m}} + r^{\frac{2}{n}}} \\ q'_s &= \frac{-\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\ q'_0 &= \frac{n-1}{n} q'_s \\ q'_1 &= \left(\frac{p}{r} \frac{1}{2n} + 1\right) q'_s + \frac{p+nq}{2r}\end{aligned}$$

For Shift, we have Equations (31), (32), (33) and (34), reproduced below, where $\eta_i = \frac{\frac{p}{n} + q}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}}$ is given by

Equation (27).

$$\begin{aligned}
\sum_{j=1}^{m-1} s'_j &= \sum_{j=2}^{m-2} s'_j = -2(m-3)\eta_2 + \sum_{i=3}^{m-3} \left(\frac{m^2 - m - 2 - 2i(2m-3) + 2i^2}{2} \right) \eta_i \\
q'_s &= \frac{-\left(\frac{p}{n} + q\right) \left(\sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} + 3 - \frac{1}{n} + \frac{q}{r} \frac{n}{2} \frac{m-1}{m} \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\
q'_0 &= \frac{n-1}{n} q'_s \\
q'_1 &= \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p+nq}{2r}
\end{aligned}$$

For both processes, we have $q'_0 = \frac{n-1}{n} q'_s$ and $q'_1 = \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p+nq}{2r}$, even though the value of q'_s will differ between the two processes. Therefore, we can use these formulas to express difference $q'_1 - q'_0$ in the relatedness formulas in terms of model parameters and q'_s only.

$$\begin{aligned}
q'_1 - q'_0 &= \left(\frac{p}{r} \frac{1}{2n} + 1 \right) q'_s + \frac{p+nq}{2r} - \frac{n-1}{n} q'_s = \left(\frac{p}{r} \frac{1}{2n} + 1 - \frac{n-1}{n} \right) q'_s + \frac{p+nq}{2r} \\
&= \left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + \frac{p+nq}{2r}
\end{aligned}$$

Now, if we plug this back into the relatedness formulas, we get the following expressions:

$$\begin{aligned}
r_0 &= \frac{(m-1) \left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i}{\left(\frac{n-1}{n} + (m-1) \left(\frac{p}{r} \frac{1}{2n} + 1 \right) \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i} \\
r_1 &= \frac{-\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s - \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i}{\left(\frac{n-1}{n} + (m-1) \left(\frac{p}{r} \frac{1}{2n} + 1 \right) \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i}
\end{aligned}$$

Using the relatedness formulas and the relationships between IBD probabilities q'_s , q'_0 and q'_1 , we can rewrite $1 - r_0$ and $r_0 - r_1$ as follows

$$\begin{aligned}
1 - r_0 &= \frac{mq'_0}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} = \frac{m \frac{n-1}{n} q'_s}{\left((m-1) \left(\frac{p}{r} \frac{1}{2n} + 1 \right) + \frac{n-1}{n} \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i} \\
r_0 - r_1 &= \frac{m(q'_1 - q'_0)}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} = \frac{m \left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + m \frac{p+nq}{2r}}{\left((m-1) \left(\frac{p}{r} \frac{1}{2n} + 1 \right) + \frac{n-1}{n} \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i}
\end{aligned}$$

If we plug these into the formulas for the critical ratios, we get, for BD,

$$\begin{aligned}
\frac{b}{c} &> \frac{p}{q} \frac{1}{n} \frac{m}{m-1} \frac{\frac{mq'_0}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i}}{\frac{m(q'_1 - q'_0)}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i}} \\
&= \frac{p}{q} \frac{1}{n} \frac{m}{m-1} \frac{q'_0}{q'_1 - q'_0} \\
&= \frac{p}{q} \frac{1}{n} \frac{m}{m-1} \frac{\frac{n-1}{n} q'_s}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + \frac{p+nq}{2r}} \\
&= \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{q'_s}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + \frac{p+nq}{2r}}
\end{aligned} \tag{38}$$

And for Shift, we get

$$\begin{aligned}
\frac{b}{c} &> \frac{p}{q} \frac{1}{n} \frac{mq'_0}{q'_0 + (m-1)q'_1 + \sum_{i=1}^{m-1} s'_i} \\
&= \frac{p}{q} \frac{1}{n} \frac{mq'_0}{(m-1)(q'_1 - q'_0) + \sum_{i=1}^{m-1} s'_i} \\
&= \frac{p}{q} \frac{1}{n} \frac{m^{\frac{n-1}{n}} q'_s}{(m-1) \left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i} \\
&= \frac{p}{q} \frac{n-1}{n^2} \frac{mq'_s}{(m-1) \left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + (m-1) \frac{p+nq}{2r} + \sum_{i=1}^{m-1} s'_i} \\
&= \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{q'_s}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) q'_s + \frac{p+nq}{2r} + \frac{1}{m-1} \sum_{i=1}^{m-1} s'_i}
\end{aligned} \tag{39}$$

What we will do in this section, is to compare these two critical b/c ratios, and show that the one for BD is always higher than the one for Shift.

We start by comparing the IBD probabilities q'_s , which now get a superscript, depending on the update process.

$$\begin{aligned}
(q'_s)^{BD} &= - \frac{\left(\frac{p}{n} + q \right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m} \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\
(q'_s)^{Shift} &= - \frac{\left(\frac{p}{n} + q \right) \left[\sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} + 3 - \frac{1}{n} + \frac{q}{r} \frac{n}{2} \frac{m-1}{m} \right]}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\
(q'_s)^{BD} - (q'_s)^{Shift} &= - \frac{\left(\frac{p}{n} + q \right) \left(m - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} - 3 \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\
(q'_s)^{BD} &= (q'_s)^{Shift} - \frac{\left(\frac{p}{n} + q \right) \left(m - 3 - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} := (q'_s)^{Shift} - A
\end{aligned} \tag{40}$$

For the terms $\frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}}$ within the sum above, we have the following calculations.

$$\begin{aligned}
\frac{i(m-i)}{m} - \frac{m-1}{m} &= \frac{i(m-i) - m + 1}{m} = \frac{im - i^2 - m + 1}{m} = \frac{m(i-1) - (i^2 - 1)}{m} = \frac{m(i-1) - (i-1)(i+1)}{m} \\
&= \frac{(i-1)(m-i-1)}{m}
\end{aligned}$$

For any $2 \leq i \leq m-2$, we have $i-1 > 0$ and $m-i-1 > 0$, which implies

$$\begin{aligned}
0 &< \frac{i(m-i)}{m} - \frac{m-1}{m} \\
\frac{m-1}{m} &< \frac{i(m-i)}{m} \\
q \frac{m-1}{m} &< q \frac{i(m-i)}{m} \\
q \frac{m-1}{m} + r \frac{2}{n} &< q \frac{i(m-i)}{m} + r \frac{2}{n} \\
\frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} &< 1
\end{aligned}$$

Therefore, every term within the summation in the numerator of Equation (40) is less than one. Hence, the sum is less than $m-3$, and the numerator is positive, which implies that A is positive. Therefore, $(q'_s)^{BD}$ is less than $(q'_s)^{Shift}$, and since we know that both q'_s 'es are negative, this implies that $(q'_s)^{BD}$ is “more negative” than $(q'_s)^{Shift}$.

To compare the critical ratios for the two processes – given below – we need to compare their last terms, as their first three terms are identical.

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{(q'_s)^{BD}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n}\right) (q'_s)^{BD} + \frac{p+nq}{2r}} \quad (41)$$

for BD, and

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{(q'_s)^{Shift}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n}\right) (q'_s)^{Shift} + \frac{p+nq}{2r} + \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift}} \quad (42)$$

for Shift.

S.5.1 When m is odd

In the calculations in this and the next subsection, we assume that $m > 3$ and $n > 1$ since these are the interesting cases. If $m \leq 3$, the two update processes become identical, and so do their thresholds. And if $n = 1$, in every individual event, the offspring replaces the parent and nothing changes in the population state.

First assume that m is odd – the case where m is even is treated below. If m is odd, we can rewrite $\sum_{i=2}^{m-2} (s'_i)^{Shift}$ as follows:

$$\begin{aligned}
\sum_{i=2}^{m-2} (s'_i)^{Shift} &= -2(m-3) \frac{\frac{p}{n} + q}{q \frac{2(m-2)}{m} + r \frac{2}{n}} + \sum_{i=3}^{m-3} \left(\frac{m^2 - m - 2 - 2i(2m-3) + 2i^2}{2} \right) \frac{\frac{p}{n} + q}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \\
&= -2(m-3) \frac{\frac{p}{n} + q}{q \frac{2(m-2)}{m} + r \frac{2}{n}} + \sum_{i=3}^{(m-1)/2} 2m \left(\frac{m-1}{m} - \frac{i(m-i)}{m} \right) \frac{\frac{p}{n} + q}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \\
&= -2(m-3) \frac{\frac{p}{n} + q}{q \frac{2(m-2)}{m} + r \frac{2}{n}} - 2m \sum_{i=3}^{(m-1)/2} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{\frac{p}{n} + q}{q \frac{i(m-i)}{m} + r \frac{2}{n}}
\end{aligned} \quad (43)$$

Going from the first to the second line, we gather the coefficients of $\frac{\frac{p}{n}+q}{q\frac{i(m-i)}{m}+r\frac{2}{n}}$ for i and $m-i$ within the sum since that is when these terms are the same. These coefficients then add up to $\frac{m^2-m-2-2i(2m-3)+2i^2}{2} + \frac{m^2-m-2-2(m-i)(2m-3)+2(m-i)^2}{2} = 2(m-1) - 2i(m-i)$.

We need to find out whether the critical ratio for the BD process is always larger than that of the Shift process; hence, we need to compare the right-hand sides of the above two equations for the critical b/c ratios, given in Equations (41) and (42). To do so, we start with comparing a few terms to zero; and step-by-step, we will make our way to the equations above. We start by showing that the four terms below are positive:

- Term 1:

$$\frac{2m}{m-1} - \frac{2m}{2m\frac{r}{qn}\left(m - \frac{1}{n}\right) + m - 1} > 0$$

since $2m\frac{r}{qn}\left(m - \frac{1}{n}\right) > 0$. Hence,

$$\frac{2m}{m-1} - \frac{\frac{qn}{r}}{m - \frac{1}{n} + \frac{qn}{2r}\frac{m-1}{m}} > 0$$

- Term 2: For $1 < i < m-1$,

$$\frac{1}{q\frac{i(m-i)}{m} + r\frac{2}{n}} > 0$$

- Term 3: As seen before,

$$\frac{i(m-i)}{m} - \frac{m-1}{m} = \frac{(i-1)(m-i-1)}{m}$$

For i ranging from 2 to $m-2$, we have $i-1 > 0$ and $m-i-1 > 0$. Therefore,

$$\frac{i(m-i)}{m} - \frac{m-1}{m} > 0$$

- Term 4:

$$\frac{2(m-3)}{m-1} - \frac{2(m-3)}{2m\frac{r}{qn}\left(m - \frac{1}{n}\right) + m - 1} > 0$$

since $2m\frac{r}{qn}\left(m - \frac{1}{n}\right) > 0$, and also

$$\frac{1}{q\frac{2(m-2)}{m} + r\frac{2}{n}} > 0$$

Hence,

$$\left(\frac{2(m-3)}{m-1} - \frac{\frac{qn}{r}\frac{m-3}{m}}{m - \frac{1}{n} + \frac{qn}{2r}\frac{m-1}{m}} \right) \frac{1}{q\frac{2(m-2)}{m} + r\frac{2}{n}} > 0$$

Since each of the terms above are individually positive for $1 < i < m-1$, their products and sums will be positive as well. If we multiply the first three terms for a given i , the resulting product will be positive. If

we sum these products over i , where $3 \leq i \leq \frac{m-1}{2}$, the resulting sum will be positive as well, since each term in the summation is positive. And finally, we add the fourth term above to the summation to reach the expression below, which is again positive:

$$0 < \left(\frac{2(m-3)}{m-1} - \frac{\frac{qn}{r} \frac{m-3}{m}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} \\ + \sum_{i=3}^{(m-1)/2} \left(\frac{2m}{m-1} - \frac{\frac{qn}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}}$$

Now, we split this expression into two parts, depending on the sign of the terms, and put the terms with a negative coefficient on the left hand side:

$$0 < \frac{2(m-3)}{m-1} \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} - \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \frac{q^{\frac{m-3}{m}}}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} \\ + \sum_{i=3}^{(m-1)/2} \left(\frac{2m}{m-1} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) - \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(q^{\frac{i(m-i)}{m}} - q^{\frac{m-1}{m}} \right) \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \\ 0 < \frac{2(m-3)}{m-1} \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} - \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \frac{q^{\frac{2(m-2)}{m}} - q^{\frac{m-1}{m}}}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} \\ + \sum_{i=3}^{(m-1)/2} \left(\frac{2m}{m-1} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) - \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(q^{\frac{i(m-i)}{m}} - q^{\frac{m-1}{m}} \right) \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \\ 0 < \frac{2(m-3)}{m-1} \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} - \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \frac{q^{\frac{2(m-2)}{m}} - q^{\frac{m-1}{m}}}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} \\ + \sum_{i=3}^{(m-1)/2} \frac{2m}{m-1} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \\ - \sum_{i=3}^{(m-1)/2} \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(q^{\frac{i(m-i)}{m}} - q^{\frac{m-1}{m}} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}}$$

This leads to

$$\frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(\sum_{i=2}^{(m-1)/2} \frac{q^{\frac{i(m-i)}{m}} - q^{\frac{m-1}{m}}}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \right) \\ < \frac{1}{m-1} \left(2(m-3) \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} + 2m \sum_{i=3}^{(m-1)/2} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \right)$$

Notice that the whole term inside the parentheses on the right hand side is the summation from Equation

(43) divided by $-\left(\frac{p}{n} + q\right)$, so we can rewrite it as $-\frac{1}{\frac{p}{n} + q} \sum_{i=2}^{m-2} (s'_i)^{Shift}$:

$$\begin{aligned}
& \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(\sum_{i=2}^{(m-1)/2} \frac{q \frac{i(m-i)}{m} - q \frac{m-1}{m}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) < \frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \left(- \sum_{i=2}^{m-2} (s'_i)^{Shift} \right) \\
& \frac{\frac{n}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(\sum_{i=2}^{(m-1)/2} \frac{q \frac{i(m-i)}{m} + r \frac{2}{n} - (q \frac{m-1}{m} + r \frac{2}{n})}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) < \frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \left(- \sum_{i=2}^{m-2} (s'_i)^{Shift} \right) \\
& \frac{\frac{2n}{2r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(\sum_{i=2}^{(m-1)/2} \left(1 - \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) \right) < \frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \left(- \sum_{i=2}^{m-2} (s'_i)^{Shift} \right) \\
& \frac{\frac{n}{2r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(\sum_{i=2}^{m-2} \left(1 - \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) \right) < \frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \left(- \sum_{i=2}^{m-2} (s'_i)^{Shift} \right) \\
& \frac{\frac{n}{2r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(m - 3 - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) < \frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \left(- \sum_{i=2}^{m-2} (s'_i)^{Shift} \right)
\end{aligned}$$

Going from the third to the fourth line above, we use that the terms for i and $m-i$ in the latter summation are always the same. Also we rewrite $\sum_{i=2}^{m-2} 1$ as $m-3$ in the last line. Now, multiply every term with $\frac{p}{n} + q$ to get;

$$\frac{\frac{p+nq}{2r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(m - 3 - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) < \frac{1}{m-1} \left(- \sum_{i=2}^{m-2} (s'_i)^{Shift} \right)$$

Now, if we multiply both sides with -1 , the sign also changes:

$$- \frac{\frac{p+nq}{2r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(m - 3 - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} \right) > \frac{1}{m-1} \left(\sum_{i=2}^{m-2} (s'_i)^{Shift} \right)$$

Another step of re-arranging the terms above gives us

$$- \frac{p+nq}{2r} \left(\frac{m - 3 - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) > \frac{1}{m-1} \left(\sum_{i=2}^{m-2} (s'_i)^{Shift} \right)$$

Multiplying and dividing the left hand side by the term $\left(\frac{p}{n} + q\right) / \left(\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}\right)$:

$$- \frac{p+nq}{2r} \left(\frac{\left(\frac{p}{n} + q\right) \left(m - 3 - \sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}}\right)}{\frac{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}}{\left(\frac{p}{n} + q\right) \left(m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}} \right) > \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift} \quad (44)$$

Notice that the term in the numerator within the parentheses on the left hand side is equal to $A = (q'_s)^{Shift} -$

$(q'_s)^{BD}$ from Equation (40), and that the denominator is equal to $-(q'_s)^{BD}$ from Equation (20).

$$\begin{aligned} -\frac{p+nq}{2r} \left(\frac{(q'_s)^{Shift} - (q'_s)^{BD}}{-(q'_s)^{BD}} \right) &> \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift} \\ \frac{p+nq}{2r} \left(\frac{(q'_s)^{Shift} - (q'_s)^{BD}}{(q'_s)^{BD}} \right) &> \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift} \end{aligned}$$

If we divide both sides by $(q'_s)^{Shift}$,

$$\frac{p+nq}{2r} \left(\frac{(q'_s)^{Shift} - (q'_s)^{BD}}{(q'_s)^{Shift}(q'_s)^{BD}} \right) < \frac{1}{m-1} \frac{1}{(q'_s)^{Shift}} \sum_{i=1}^{m-1} (s'_i)^{Shift}$$

where the sign of the inequality changes since we multiply both sides with a negative term. Using $\frac{(q'_s)^{Shift} - (q'_s)^{BD}}{(q'_s)^{Shift}(q'_s)^{BD}} = \frac{1}{(q'_s)^{BD}} - \frac{1}{(q'_s)^{Shift}}$, we can rewrite the above inequality as follows,

$$\begin{aligned} \frac{p+nq}{2r} \left(\frac{1}{(q'_s)^{BD}} - \frac{1}{(q'_s)^{Shift}} \right) &< \frac{1}{m-1} \frac{1}{(q'_s)^{Shift}} \sum_{i=1}^{m-1} (s'_i)^{Shift} \\ \frac{p+nq}{2r} \frac{1}{(q'_s)^{BD}} &< \frac{p+nq}{2r} \frac{1}{(q'_s)^{Shift}} + \frac{1}{m-1} \frac{1}{(q'_s)^{Shift}} \sum_{i=1}^{m-1} (s'_i)^{Shift} \end{aligned}$$

Add $\frac{p}{r} \frac{1}{2n} + \frac{1}{n}$ to both sides

$$\frac{p}{r} \frac{1}{2n} + \frac{1}{n} + \frac{p+nq}{2r} \frac{1}{(q'_s)^{BD}} < \frac{p}{r} \frac{1}{2n} + \frac{1}{n} + \frac{p+nq}{2r} \frac{1}{(q'_s)^{Shift}} + \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift} \frac{1}{(q'_s)^{Shift}}$$

Reverse the numerator and the denominator on both sides of the inequality,

$$\begin{aligned} \frac{1}{\frac{p}{r} \frac{1}{2n} + \frac{1}{n} + \frac{p+nq}{2r} \frac{1}{(q'_s)^{BD}}} &> \frac{1}{\frac{p}{r} \frac{1}{2n} + \frac{1}{n} + \frac{p+nq}{2r} \frac{1}{(q'_s)^{Shift}} + \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift} \frac{1}{(q'_s)^{Shift}}} \\ \frac{(q'_s)^{BD}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) (q'_s)^{BD} + \frac{p+nq}{2r}} &> \frac{(q'_s)^{Shift}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) (q'_s)^{Shift} + \frac{p+nq}{2r} + \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift}} \end{aligned}$$

where the sign of the inequality changes again on the first line above since we are reversing the fractions that we are comparing. Now, if we multiply both sides by $\frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1}$, we arrive at the inequality:

$$\begin{aligned} \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{(q'_s)^{BD}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) (q'_s)^{BD} + \frac{p+nq}{2r}} &> \\ \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{(q'_s)^{Shift}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n} \right) (q'_s)^{Shift} + \frac{p+nq}{2r} + \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift}} & \end{aligned}$$

In this last inequality, the left hand side is the critical b/c ratio for the BD process, and the right hand side is the critical b/c ratio for the Shift process, given in Equations (41) and (42), respectively.

S.5.2 When m is even

If m is even, we can rewrite $\sum_{i=2}^{m-2} (s'_i)^{Shift}$ as follows:

$$\begin{aligned}
\sum_{i=2}^{m-2} (s'_i)^{Shift} &= -2(m-3) \frac{\frac{p}{n} + q}{q^{\frac{2(m-2)}{m}} + r \frac{2}{n}} + \sum_{i=3}^{m-3} \left(\frac{m^2 - m - 2 - 2i(2m-3) + 2i^2}{2} \right) \frac{\frac{p}{n} + q}{q^{\frac{i(m-i)}{m}} + r \frac{2}{n}} \\
&= -2(m-3) \frac{\frac{p}{n} + q}{q^{\frac{2(m-2)}{m}} + r \frac{2}{n}} - 2m \sum_{i=3}^{m/2-1} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{\frac{p}{n} + q}{q^{\frac{i(m-i)}{m}} + r \frac{2}{n}} \\
&\quad - \frac{(m-2)^2}{4} \frac{\frac{p}{n} + q}{q^{\frac{\frac{m}{2} \cdot \frac{m}{2}}{m}} + r \frac{2}{n}}
\end{aligned} \tag{45}$$

This we can do for reasons similar to the ones when m is odd.

To compare the critical ratios for the two update processes from Equations (41) and (42), we are going to follow a very similar path to the case above in Section S.5.1. We start with adding a fifth term to the terms given in the previous subsection; and step-by-step, we are going to reach the equations for the critical ratios of the two update processes. We have shown previously that the Terms 1 through 4 are positive. Here, we add another term and show that it is also positive:

- Term 5:

$$\frac{(m-2)^2}{4(m-1)} - \frac{(m-2)^2}{8m \frac{r}{qn} \left(m - \frac{1}{n}\right) + 4(m-1)} > 0$$

since $8m \frac{r}{qn} \left(m - \frac{1}{n}\right) > 0$, therefore,

$$\frac{(m-2)^2}{4(m-1)} - \frac{\frac{qn}{r} \frac{(m-2)^2}{8m}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} > 0$$

and

$$\frac{1}{q^{\frac{m^2}{4m}} + r \frac{2}{n}} > 0$$

Hence,

$$\left(\frac{(m-2)^2}{4(m-1)} - \frac{\frac{qn}{r} \frac{(m-2)^2}{8m}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) \frac{1}{q^{\frac{m^2}{4m}} + r \frac{2}{n}} > 0$$

Now, if we multiply Term 1, 2 and 3 and sum these products over $3 \leq i \leq \frac{m-2}{2} = \frac{m}{2} - 1$, add Term 4 to the summation from the previous subsection, we know that the resulting term will be positive. Now, we add

Term 5 above, which is also positive, to get

$$\begin{aligned}
0 &< \left(\frac{2(m-3)}{m-1} - \frac{\frac{qn}{r} \frac{m-3}{m}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} \\
&+ \sum_{i=3}^{m/2-1} \left(\frac{2m}{m-1} - \frac{\frac{qn}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \\
&+ \left(\frac{(m-2)^2}{4(m-1)} - \frac{\frac{qn}{r} \frac{(m-2)^2}{8m}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \right) \frac{1}{q^{\frac{m^2}{4m}} + r^{\frac{2}{n}}}
\end{aligned}$$

We are re-arranging the expression above such that every term with a negative (positive) sign is on the left hand side (right hand side),

$$\begin{aligned}
&\frac{\frac{qn}{r} \left(\frac{2(m-2)}{m} - \frac{m-1}{m} \right)}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} + \sum_{i=3}^{m/2-1} \frac{\frac{qn}{r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \\
&+ \frac{\frac{qn}{r} \frac{(m-2)^2}{8m}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \frac{1}{q^{\frac{m^2}{4m}} + r^{\frac{2}{n}}} \\
&< \frac{2(m-3)}{m-1} \frac{1}{q^{\frac{2(m-2)}{m}} + r^{\frac{2}{n}}} + \sum_{i=3}^{m/2-1} \frac{2m}{m-1} \left(\frac{i(m-i)}{m} - \frac{m-1}{m} \right) \frac{1}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \\
&+ \frac{(m-2)^2}{4(m-1)} \frac{1}{q^{\frac{m^2}{4m}} + r^{\frac{2}{n}}}
\end{aligned}$$

Here, we used $\frac{2(m-2)}{m} - \frac{m-1}{m} = \frac{m-3}{m}$ on the first line. Realize that the right hand side in the last line above is equal to $-\frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \sum_{i=2}^{m-2} (s'_i)^{Shift}$ where $\sum_{i=2}^{m-2} (s'_i)^{Shift}$ is defined in Equation (45).

Using similar steps as in the previous subsection and the fact that

$$\frac{\frac{m}{2} \left(m - \frac{m}{2} \right)}{m} - \frac{m-1}{m} = \frac{m^2}{4m} - \frac{m-1}{m} = \frac{(m-2)^2}{4m}$$

and

$$\frac{q^{\frac{m^2}{4m}} - q^{\frac{m-1}{m}}}{q^{\frac{m^2}{4m}} + r^{\frac{2}{n}}} = 1 - \frac{q^{\frac{m-1}{m}} + r^{\frac{2}{n}}}{q^{\frac{m^2}{4m}} + r^{\frac{2}{n}}}$$

we get

$$\frac{\frac{n}{2r}}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(m - 3 - \sum_{i=2}^{m-2} \frac{q^{\frac{m-1}{m}} + \frac{2r}{n}}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \right) < -\frac{1}{m-1} \frac{1}{\frac{p}{n} + q} \sum_{i=2}^{m-2} (s'_i)^{Shift}$$

Multiplying both sides by $-\left(\frac{p}{n} + q\right)$, which changes the sign of the inequality, and multiplying and dividing the left hand side by $\left(\frac{p}{n} + q\right) / \left(\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}\right)$ gives

$$-\frac{p+nq}{2r} \frac{1}{m - \frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}} \left(m - 3 - \sum_{i=2}^{m-2} \frac{q^{\frac{m-1}{m}} + \frac{2r}{n}}{q^{\frac{i(m-i)}{m}} + r^{\frac{2}{n}}} \right) > \frac{1}{m-1} \sum_{i=2}^{m-2} (s'_i)^{Shift}$$

$$\frac{p+nq}{2r} \left(\frac{\left(\frac{p}{n}+q\right) \left(m-3-\sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}}\right)}{\frac{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} - \frac{\left(\frac{p}{n}+q\right) \left(m-\frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\frac{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}}} \right) > \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift}$$

which is the same inequality as in Equation (44). Following the same steps after Equation (44) in the previous subsection, we arrive at the result that

$$\frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{(q'_s)^{BD}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n}\right) (q'_s)^{BD} + \frac{p+nq}{2r}} > \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{(q'_s)^{Shift}}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n}\right) (q'_s)^{Shift} + \frac{p+nq}{2r} + \frac{1}{m-1} \sum_{i=1}^{m-1} (s'_i)^{Shift}}$$

In this last inequality, the left hand side is the critical b/c ratio for the BD process, and the right hand side is the critical b/c ratio for the Shift process, given in Equations (41) and (42), respectively.

S.5.3 Limit results for the number of groups approaching infinity

In this section, we explore the results in the limit where the number of groups m approaches infinity.

S.5.3.1 BD

For the BD process, we first repeat and rewrite Equation (38)

$$\begin{aligned} \frac{b}{c} &> \frac{p}{q} \frac{1}{n} \frac{m}{m-1} \frac{\frac{n-1}{n} q'_s}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n}\right) q'_s + \frac{p+nq}{2r}} \\ &= \frac{p}{q} \frac{1}{n} \frac{m}{m-1} \frac{\frac{n-1}{n}}{\frac{p}{r} \frac{1}{2n} + \frac{1}{n} + \frac{p+nq}{2r} \frac{1}{q'_s}} \end{aligned}$$

as well as Equation (20)

$$q'_s = \frac{-\left(\frac{p}{n}+q\right) \left(m-\frac{1}{n} + \frac{qn}{2r} \frac{m-1}{m}\right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}}$$

Since $\lim_{m \rightarrow \infty} q'_s = \infty$, the critical b/c ratio for the BD process converges to

$$\begin{aligned} \frac{b}{c} &> \frac{p}{q} \frac{1}{n} \frac{\frac{n-1}{n}}{\frac{p}{r} \frac{1}{2n} + \frac{1}{n}} \\ &= \frac{p}{q} \frac{n-1}{n} \frac{1}{\frac{p}{2r} + 1} \end{aligned} \tag{46}$$

The limit result of the critical ratio we found in Equation (46) is in line with numerical solutions for large m .

S.5.3.2 Shift

For the Shift process, we first repeat and rewrite Equation (39)

$$\begin{aligned} \frac{b}{c} &> \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{q'_s}{\left(\frac{p}{r} \frac{1}{2n} + \frac{1}{n}\right) q'_s + \frac{p+nq}{2r} + \frac{1}{m-1} \sum_{i=1}^{m-1} s'_i} \\ &= \frac{p}{q} \frac{n-1}{n^2} \frac{m}{m-1} \frac{1}{\frac{p}{r} \frac{1}{2n} + \frac{1}{n} + \frac{p+nq}{2r} \frac{1}{q'_s} + \frac{1}{m-1} \frac{1}{q'_s} \sum_{i=1}^{m-1} s'_i} \end{aligned}$$

We also have, from Equations (32) and (31), reproduced below, where $\eta_i = \frac{\frac{p}{n} + q}{q \frac{i(m-i)}{m} + r \frac{2}{n}}$ is given by Equation (27).

$$\begin{aligned} q'_s &= \frac{-\left(\frac{p}{n} + q\right) \left(\sum_{i=2}^{m-2} \frac{q \frac{m-1}{m} + r \frac{2}{n}}{q \frac{i(m-i)}{m} + r \frac{2}{n}} + 3 - \frac{1}{n} + \frac{q}{r} \frac{n}{2} \frac{m-1}{m}\right)}{\frac{pq}{r} \frac{1}{2n} \frac{m-1}{m} + q \frac{1}{n} \frac{m-1}{m} + p \frac{n-1}{n^3}} \\ \sum_{j=1}^{m-1} s'_j &= -2(m-3)\eta_2 + \sum_{i=3}^{m-3} \left(\frac{m^2 - m - 2 - 2i(2m-3) + 2i^2}{2}\right) \eta_i \end{aligned}$$

Now consider the sum in the numerator in q'_s . Since each term in this summation is less than 1, we have

$$\sum_{i=2}^{m-2} \frac{m-1 + m \frac{r}{q} \frac{2}{n}}{i(m-i) + m \frac{r}{q} \frac{2}{n}} < m-3$$

and hence q'_s decreases at most proportional to m . On the other hand, $\sum_{j=1}^{m-1} s'_j$ decreases proportional to m^2 . Therefore, the term with $\sum_{j=1}^{m-1} s'_j$ in the denominator of the critical b/c ratio for the Shift process dominates in the limit $m \rightarrow \infty$, and hence, the critical b/c ratio for Shift converges to

$$\frac{b}{c} > 0 \quad (47)$$

The limit result of the critical ratio we found in Equation (47) is in line with numerical solutions for large m .

S.5.3.3 Birth-Death versus Shift in the limit where number of groups approach infinity

In the previous subsection, we solved for the critical b/c ratios in the limit where the number of groups m approaches infinity. The limit results are repeated below for convenience.

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{1}{\frac{p}{2r} + 1}$$

for BD, and

$$\frac{b}{c} > 0$$

for Shift. From these formulas, it is immediately clear that the critical b/c ratio for Shift is lower than the one for BD in the limit $m \rightarrow \infty$.

S.6 Theoretical results and simulations

In this section, we take the critical b/c ratios in the limit of weak selection, calculated in Section S.3, combine them, as we did in Section S.5, with the relatednesses in the limit of weak selection, calculated in Section S.4, and plot them for a variety of parameter combinations. We combine those analytical results in the limit of weak selection with simulation results not in the limit of weak selection. As we vary group size n , the number of groups m , and migration rate r , we want to choose the probabilities of group versus individual events such that the ratio of probabilities for an individual to die in an individual event and in a group event remains constant under neutral selection. In this case, we choose them so that these probabilities are always equally large. At neutrality, the probability that an individual dies in an individual event is $p \frac{1}{m} \frac{1}{n}$ and the probability that an individual dies in a group event is $q \frac{1}{m}$. Keeping them equal therefore requires

$$p \frac{1}{m} \frac{1}{n} = q \frac{1}{m}$$

$$p = nq$$

We also have the condition $p + q + r = 1$. Together with the equality above, this implies that

$$p + q = 1 - r$$

$$(n + 1)q = 1 - r$$

$$q = (1 - r) \frac{1}{n + 1} \rightarrow p = (1 - r) \frac{n}{n + 1}$$

In Figures S.6 and S.7, we choose $r = 0.1$, and in Figure S.8, r varies.

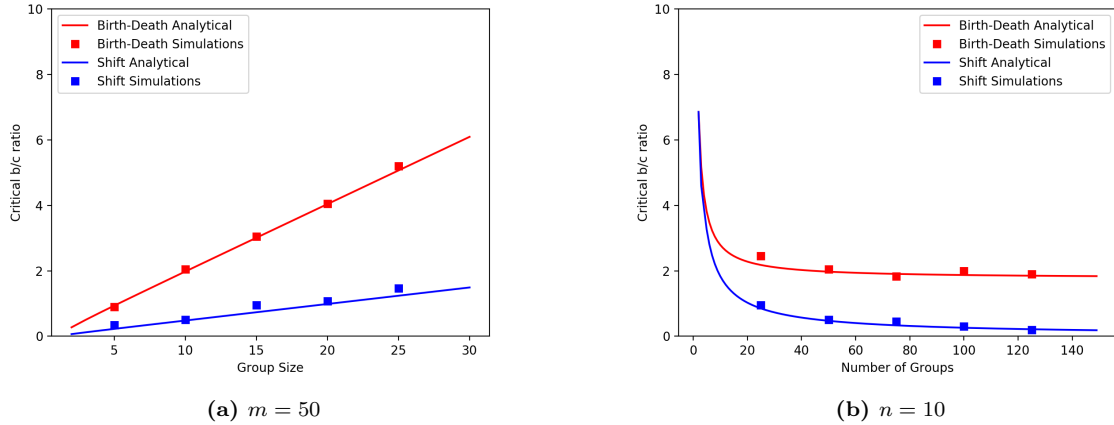


Figure S.6: Results for the critical b/c ratios, combined with simulations with 1000000 independent runs, where $w_{ind} = w_{gr} = 0.1$, and $r = 0.1$, $p = 0.9 \frac{1}{n+1}$ and $q = 0.9 \frac{n}{n+1}$. In (a), the effect of increasing the group size is shown for a fixed number of groups. In (b), the effect of increasing the number of groups is shown for a fixed group size.

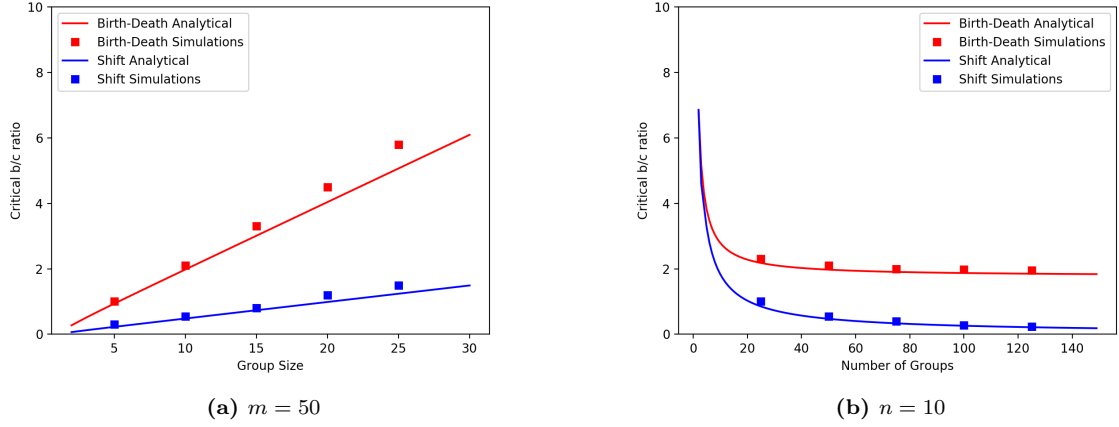


Figure S.7: Same as Figure 6, but with $w_{ind} = w_{gr} = 0.5$.

In Figures S.6a and S.7a, we see that the thresholds in both processes increase with group size, and that the gap between the critical b/c ratios for the two processes is there for a range of group sizes. In Figures S.6b and S.7b, we see that the thresholds decrease with the number of groups, and, again, that the gap between the critical b/c ratios for the two processes is there for a range of numbers of groups. Both the increase of the thresholds with group size and the decrease with the number of groups are in line with the results of Traulsen and Nowak (2006). The gap between the thresholds for the two processes in relative terms becomes (very) large for (very) large numbers of groups.

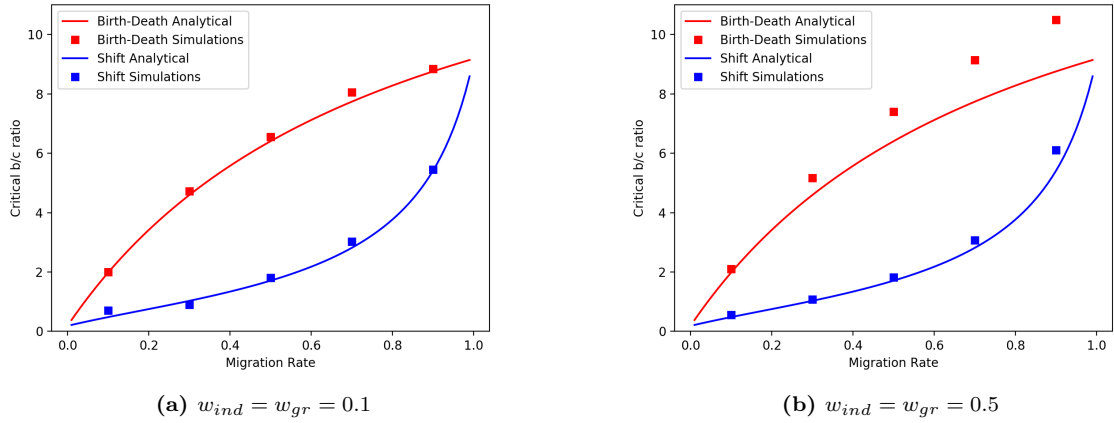


Figure S.8: Results for the critical b/c ratios for different migration rates r , for $n = 10, m = 50$, again combined with simulations with 1000000 independent runs.

In Figure S.8, we see that both thresholds increase with the migration rate. This is understandable, since a higher migration rate will result in lower relatedness. We also see that the gap disappears for migration rates close to 0 and migration rates close to 1. For migration rates close to 0, and not too few groups, there is a fair chance that the process will spend a lot of time in population states where all groups are at within-group fixation, or, in other words, where all groups consist of only cooperators or only defectors. Once this has happened, there is a fair chance that the population will get to a state where these all-cooperator groups form

one sequence of groups on the circle. Without mutation or migration, this will remain to be true from then onwards. In Section S.7, we will see that if we imagine a mutant group, where all individuals are cooperators, the fixation probability of such a mutant group is the same in either process. If cooperators on average gain ground on (or lose ground to) defectors, they do so faster with Shift than with BD, but the condition under which they gain or lose is the same for both processes. The reason for the difference in speed is that with Shift, in all intermediate population states with a string of all-cooperator groups and a string of all-defector groups, any all-cooperator group has a real chance of replacing any all-defector group and vice versa. On the other hand, in BD, all the action is at the two boundaries between the strings of all-cooperator groups and all-defector groups, while group reproduction events not on the boundary are inconsequential for the population state. This implies that the boundary moves much more with Shift than it does with BD, and if the boundary moves in expectation in favour of the all-cooperator groups, it will move faster with Shift than with BD. However, the speed turns out not to matter for the fixation probability at the group level of a mutant all-cooperator group. The speed does matter if there are mixed groups because if there are mixed groups, the ground gained at the group level will balance against the ground lost at the individual level. With cooperative groups winning faster in Shift, they can therefore overcome a larger within-group decay of cooperators. Absent mixed groups, however, speed does not matter anymore, and the difference in fixation probabilities disappears.

At migration rates r close to 1, it is not hard to understand why the difference between BD and Shift disappears as well. When almost all events are migration events, the population is shaken and stirred between any two reproduction events, be it at the individual or at the group level, and all that matters is what being a cooperator does to an individual's own birth rate and what it does to its group's birth rate. Those effects are the same in both processes, and therefore the difference between the two processes should disappear at very high migration rates.

This is also reflected in the formulas for the thresholds. In the limit of $r \uparrow 1$, only finite population effects remain, and relatednesses become $r_s = r_1 = -\frac{1}{nm-1}$. The thresholds then become:

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{m}{m-1} \frac{1-r_s}{1+(n-1)r_s - nr_1} = \frac{p}{q} \frac{n-1}{n} \frac{m}{m-1}$$

for BD, and

$$\frac{b}{c} > \frac{p}{q} \frac{n-1}{n} \frac{1-r_s}{1+(n-1)r_s} = \frac{p}{q} \frac{n-1}{n} \frac{1+\frac{1}{nm-1}}{1-(n-1)\frac{1}{nm-1}} = \frac{p}{q} \frac{n-1}{n} \frac{nm-1+1}{nm-1-n+1} = \frac{p}{q} \frac{n-1}{n} \frac{m}{m-1}$$

for Shift, and these thresholds are the same.

S.7 Analytical solutions, without migration, and for limit cases that are not the limit of weak selection

There are some limit cases, other than the limit of weak selection, for which it also becomes feasible to derive analytical solutions for fixation probabilities. These results will help understand why the difference between the critical b/c ratios for BD and for Shift disappears when migration rates get close to 0. In both limits we consider here, we assume that there is no migration ($r = 0$), we assume strong selection ($w_{ind} = w_{gr} = 1$), and we assume that there is a separation of timescales. In the first limit, where $p \rightarrow 1$, group events are rare, and in almost all events, an individual reproduces. In the second limit, where $p \rightarrow 0$, individual events are rare, and in almost all events, a group reproduces.

S.7.1 Fixation probabilities of mutant groups

To compute fixation probabilities in both limits, it will be useful first to compute the fixation probability of a “mutant group”, with l cooperators in it, in a population of groups, all of which contain k cooperators. This is the process at the group selection time scale³. It will only visit states in which there is one string of subsequent groups with k cooperators, and one string of subsequent groups with l cooperators – besides of course the absorbing states, where all groups have k cooperators or all groups have l cooperators. Therefore, the population state can be denoted by a single number i , $0 \leq i \leq m$, representing the number of groups with l cooperators.

Birth-Death

For the Birth-Death process, the only way in which i can change is if one of the groups on the two boundaries between the strings of different group types is chosen for reproduction; one of the two k -groups on the boundary if i is to go down, and one of the two l -groups on the boundary if i is to go up. Moreover, it needs to replace the group on the other side of the boundary, and not its other neighbour or itself, which happens with probability $\frac{1}{2} \frac{m-1}{m}$. The probability of a transition from i to $i-1$ in the Birth-Death process then becomes $T_{i,BD}^- = 2 \cdot \frac{1}{2} \frac{m-1}{m} \cdot \frac{1 + \frac{k}{n}b}{i(1 + \frac{l}{n}b) + (m-i)(1 + \frac{k}{n}b)}$, while the probability of a transition from i to $i+1$ is $T_{i,BD}^+ = 2 \cdot \frac{1}{2} \frac{m-1}{m} \cdot \frac{1 + \frac{l}{n}b}{i(1 + \frac{l}{n}b) + (m-i)(1 + \frac{k}{n}b)}$. That makes the down-up ratio for the Birth-Death process equal to

$$\frac{T_{i,BD}^-}{T_{i,BD}^+} = \frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b}$$

Shift

For the Shift process, change is much more likely. State i changes any time the group that is chosen for reproduction is of a different type than the group that is chosen for death. The probability of a transition from i to $i-1$ is $T_{i,Shift}^- = \frac{(m-i)(1 + \frac{k}{n}b)}{i(1 + \frac{l}{n}b) + (m-i)(1 + \frac{k}{n}b)} \frac{i}{m-1}$, while the probability of a transition from i to $i+1$ is $T_{i,Shift}^+ = \frac{i(1 + \frac{l}{n}b)}{i(1 + \frac{l}{n}b) + (m-i)(1 + \frac{k}{n}b)} \frac{m-i}{m-1}$. That makes the down-up ratio for the Shift process also equal to

$$\frac{T_{i,Shift}^-}{T_{i,Shift}^+} = \frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b}$$

³The fixation probabilities follow more or less directly from the fixation probabilities in the standard Birth-Death process on the cycle (Ohtsuki and Nowak, 2006; Nowak, 2006), or the standard Shift process (Allen and Nowak, 2012). The only difference is that in the standard versions, there are individuals on a cycle, and here we have groups on a cycle instead.

Birth-Death and Shift

Since the down-up ratios are the same for both, the fixation probability of a group with l cooperators, while all other groups have k of them, is also the same for Birth-Death and Shift:

$$\tau_{k \rightarrow l} = \frac{1}{1 + \sum_{j=1}^{m-1} \prod_{i=1}^j \frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b}} = \frac{1}{1 + \sum_{j=1}^{m-1} \left(\frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b} \right)^j} = \frac{1}{1 + \frac{1 - \left(\frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b} \right)^m}{1 - \frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b}} - 1} = \frac{1 - \frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b}}{1 - \left(\frac{1 + \frac{k}{n}b}{1 + \frac{l}{n}b} \right)^m}$$

This can be rewritten as

$$\tau_{k \rightarrow l} = \frac{\left(1 + \frac{l}{n}b\right) - \left(1 + \frac{k}{n}b\right)}{\left(1 + \frac{l}{n}b\right)^m - \left(1 + \frac{k}{n}b\right)^m} \frac{\left(1 + \frac{l}{n}b\right)^m}{1 + \frac{l}{n}b} = \frac{\frac{l-k}{n}b \left(1 + \frac{l}{n}b\right)^{m-1}}{\left(1 + \frac{l}{n}b\right)^m - \left(1 + \frac{k}{n}b\right)^m}$$

In the limit of $p \rightarrow 1$, we will use $\tau_{0 \rightarrow n}$ and $\tau_{n \rightarrow 0}$, and in the limit of $p \rightarrow 0$, we will use $\tau_{k \rightarrow k+1}$ and $\tau_{k \rightarrow k-1}$ for $1 \leq k \leq m-1$.

S.7.2 The limit $p \rightarrow 1$

In this limit, group events are rare, while there are individual events almost all of the time. This implies that within her group, an initial mutant will either have gone extinct, or gone to fixation, before a group event happens. With the separation of timescales, we can first concentrate on the probabilities of fixation within the group happening. With j being the number of cooperators in the group, a defector replaces a cooperator within the group, and a cooperator replaces a defector, with probabilities

$$T_j^- = \frac{n-j}{n-jc} \frac{j}{n} \quad \text{and} \quad T_j^+ = \frac{j(1-c)}{n-jc} \frac{n-j}{n}$$

The down-up ratio therefore is $\frac{T_j^-}{T_j^+} = \frac{1}{1-c}$, making the fixation probabilities within the group

$$\sigma_C = \frac{1 - \frac{1}{1-c}}{1 - \left(\frac{1}{1-c}\right)^n} = \frac{c(1-c)^{n-1}}{1 - (1-c)^n} \quad \text{and} \quad \sigma_D = \left(\frac{1}{1-c}\right)^{n-1} \frac{1 - \frac{1}{1-c}}{1 - \left(\frac{1}{1-c}\right)^n} = \frac{c}{1 - (1-c)^n}$$

After fixation within the group, all groups will be homogeneous. As soon as the population consists of homogeneous groups, the population state can only change with a group event. Therefore, the fixation probability of a mutant will be the product of the fixation probability of the mutant within the group – which is the same for both processes, because this does not involve group events – and the fixation probability of the all-mutant group in the population – which is also the same for both processes, as we have seen above.

$$\rho_C = \sigma_C \tau_{0 \rightarrow n} = \frac{1 - \frac{1}{1-c}}{1 - \left(\frac{1}{1-c}\right)^n} \frac{1 - \frac{1}{1+b}}{1 - \left(\frac{1}{1+b}\right)^m} = \frac{c(1-c)^{n-1}}{1 - (1-c)^n} \frac{b(1+b)^{m-1}}{(1+b)^m - 1}$$

and

$$\rho_D = \sigma_D \tau_{n \rightarrow 0} = \frac{1 - (1-c)}{1 - (1-c)^n} \frac{1 - (1+b)}{1 - (1+b)^m} = \frac{c}{1 - (1-c)^n} \frac{b}{1 - (1+b)^m}$$

S.7.3 The limit $p \rightarrow 0$

In this limit, individual events are rare, while there are group events almost all of the time. The population will therefore regularly be in a state in which all groups have the same composition. When each group has the same number of cooperators, the population state can only change as the result of an individual event. If it does, then that event can make the number of cooperators within one group go up by one, or go down by one. After that, a sequence of group events either make the new “mutant group” go extinct, or go to fixation. Fixation of the mutant group happens with probability $\tau_{k \rightarrow k+1}$ if the individual event had a cooperator replace a defector, and with probability $\tau_{k \rightarrow k-1}$ if a defector replaced a cooperator. With rare individual events, the population, on the larger time scale, therefore moves between states where all groups have the same number of cooperators, and this number goes up or down by at most 1. The probability that it goes down by one is the probability with which a defector reproduces, which is $\frac{n-k}{n-kc}$, times the probability that a cooperator dies, which is $\frac{k}{n}$, times the fixation probability of the “mutant group”, which is $\tau_{k \rightarrow k-1}$. The probability that it goes up by one is the probability with which a cooperator reproduces, which is $\frac{k(1-c)}{n-kc}$, times the probability that a defector dies, which is $\frac{n-k}{n}$, times the fixation probability of the “mutant group”, which is $\tau_{k \rightarrow k+1}$. The down-up ratio at the larger timescale therefore is

$$\frac{\mathbf{T}_k^-}{\mathbf{T}_k^+} = \frac{1}{1-c} \frac{\tau_{k \rightarrow k-1}}{\tau_{k \rightarrow k+1}}$$

With these down-up ratios at the larger timescale, we find the overall fixation probabilities of cooperators and defectors.

$$\lim_{p \rightarrow 0} \rho_C = \tau_{0 \rightarrow 1} \frac{1}{1 + \sum_{k=1}^{n-1} \prod_{l=1}^k \frac{\mathbf{T}_l^-}{\mathbf{T}_l^+}} = \tau_{0 \rightarrow 1} \frac{1}{1 + \sum_{k=1}^{n-1} \left(\frac{1}{1-c} \right)^k \prod_{l=1}^k \frac{\tau_{l \rightarrow l-1}}{\tau_{l \rightarrow l+1}}}$$

and

$$\lim_{p \rightarrow 0} \rho_D = \tau_{n \rightarrow n-1} \frac{\prod_{k=1}^{n-1} \frac{\mathbf{T}_k^-}{\mathbf{T}_k^+}}{1 + \sum_{k=1}^{n-1} \prod_{l=1}^k \frac{\mathbf{T}_l^-}{\mathbf{T}_l^+}} = \tau_{n \rightarrow n-1} \frac{\left(\frac{1}{1-c} \right)^{n-1} \prod_{l=1}^{n-1} \frac{\tau_{l \rightarrow l-1}}{\tau_{l \rightarrow l+1}}}{1 + \sum_{k=1}^{n-1} \left(\frac{1}{1-c} \right)^k \prod_{l=1}^k \frac{\tau_{l \rightarrow l-1}}{\tau_{l \rightarrow l+1}}}$$

Finally, one can further rewrite the product of the ratio of τ 's in these formulas as follows.

$$\prod_{l=1}^k \frac{\tau_{l \rightarrow l-1}}{\tau_{l \rightarrow l+1}} = \frac{1}{\left(1 + \frac{k}{n}b\right)^{m-1}} \frac{\left(1 + \frac{1}{n}b\right)^{m-1}}{\left(1 + \frac{k+1}{n}b\right)^{m-1}} \frac{\left(1 + \frac{k+1}{n}b\right)^m - \left(1 + \frac{k}{n}b\right)^m}{\left(1 + \frac{1}{n}b\right)^m - 1}$$

The down-up ratios can be smaller or larger than 1, and the fixation probability could be below or above $1/(mn)$, depending on the parameters.

The probability includes the effect of both individual and group events. Defectors have a higher fixation probability within the group. However, cooperator groups have a higher fixation probability among groups. The overall comparison again depends on the parameter choices.

S.8 Empirical implications

S.8.1 F_{ST}

Many empirical papers estimate F_{ST} 's. Our definition of r_s follows the definition of relatedness in Rousset (2004) and Durrett (2008), applied to members of the same group, and our definition of relatedness between members of different groups applies the same definition to individuals from different groups. Many empiricists however use other definitions. We assume that the definitions empiricists use are equivalent to the definition from Rousset (2004) and Durrett (2008), applied to members of the same group, but we have not found a reference where this is proven formally. What we will do below is to indicate how the two definitions used by most empiricists are the same as the probabilistic definition of relatedness within the group that we also used at the end of Section S.4. There, relatedness between two individuals in the same group, for a given population state k , is defined as $r_{s,k} = \mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D)$, where the probabilities refer to the outcomes of two subsequent draws from the same group, without replacement. Relatedness r_s is then defined as the weighted average $r_s = \sum_k q_k r_{s,k}$ of those relatednesses over different states, where the weights q_k of states are given by the invariant distribution of the Markov chain in the limit of no mutation, normalized after excluding the absorbing states where all individuals are cooperators, or all are defectors (Allen and Tarnita, 2014).

Our model has groups of equal size, and let f_i be the fraction of groups with i cooperators out of n group members in population state k (we suppress the population state k in the notation). The overall frequency of cooperators and defectors follow from these fractions of group types: $\sum_{i=0}^n f_i \frac{i}{n} = p$ and $\sum_{i=0}^n f_i \frac{n-i}{n} = 1 - p$. Also the fractions of group types must add up to 1; $\sum_{i=0}^n f_i = 1$. Then, we can rewrite $r_{s,k}$ as

$$\begin{aligned}
 r_{s,k} &= \mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D) = \frac{\sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1}}{\sum_{i=0}^n f_i \frac{i}{n}} - \frac{\sum_{i=0}^n f_i \frac{n-i}{n} \frac{i}{n-1}}{\sum_{i=0}^n f_i \frac{n-i}{n}} = \frac{\sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1}}{p} - \frac{\sum_{i=0}^n f_i \frac{n-i}{n} \frac{i}{n-1}}{1-p} \\
 &= \frac{\sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} (1-p)}{p(1-p)} - \frac{\sum_{i=0}^n f_i \frac{n-i}{n} \frac{i}{n-1} p}{p(1-p)} \\
 &= \frac{\sum_{i=0}^n f_i \frac{i}{n} \left(1 - \frac{n-i}{n-1}\right) (1-p)}{p(1-p)} - \frac{\sum_{i=0}^n f_i \frac{n-i}{n} \frac{i}{n-1} p}{p(1-p)} \\
 &= \frac{\sum_{i=0}^n f_i \frac{i}{n} (1-p) - \sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1} + \sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1} p}{p(1-p)} - \frac{\sum_{i=0}^n f_i \frac{n-i}{n} \frac{i}{n-1} p}{p(1-p)} \\
 &= \frac{p(1-p) - \sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1}}{p(1-p)}
 \end{aligned}$$

The final expression matches the way F_{ST} is regularly defined (see Equation 48), except for one detail. With groups numbered from $j = 1, \dots, m$, the normal definition would have $\sum_j^m p_j(1-p_j)$ as a second term in the numerator, where $p_j(1-p_j)$ is the variance within group j , and $p(1-p)$ is the variance in the population as a whole. In our case, with f_i counting the numbers of groups with i cooperators out of n group members, this average within group variance would be $\sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n}$. In our formula, we have $\sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1}$ instead.

Here it is worth noting that the variance is half the expected squared difference between two independent draws from the same distribution. If X_1 and X_2 are two independent draws for the same random variable (which therefore have the same first and second moment), then the expected value of the squared difference

between them is

$$\mathbb{E}[X_1 - X_2]^2 = \mathbb{E}[X_1^2 - X_2^2 - 2X_1X_2] = 2\mathbb{E}[X_1^2] - \mathbb{E}^2[X_1] = 2\text{Var}(X_1)$$

In our case, X_1 and X_2 represent drawing an individual from one and the same group, where drawing a cooperator makes the value of X to be 1, and drawing a defector makes it 0. If we draw two individuals without replacement, instead of with replacement, then they are no longer independent. We can however still compute the expected squared difference between the draws. This is

$$\frac{i}{n} \frac{i-1}{n-1} \cdot 0^2 + \frac{i}{n} \frac{n-i}{n-1} \cdot 1^2 + \frac{n-i}{n} \frac{i}{n-1} \cdot 1^2 + \frac{n-i}{n} \frac{n-i-1}{n-1} \cdot 0^2 = 2 \frac{i(n-i)}{n(n-1)}$$

Half of this is what is used instead of the within-group variance. For every state k , the "without-replacement" version of the F_{ST} is therefore equivalent to the definition as a difference in conditional probabilities $r_{s,k} = \mathbb{P}_{s,k}(C|C) - \mathbb{P}_{s,k}(C|D)$. For large group sizes n , the difference between the with and without replacement version of the within-group variance disappears.

$$F_{ST} = \frac{\text{population variance} - \text{average within group variance}}{\text{population variance}} \quad (48)$$

This definition of the F_{ST} features both the average within-group variance $\sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1}$ (with replacement) and the overall variance. There is yet another definition of F_{ST} , which expresses it in terms of within-group and between-group variance. We can take the following steps to get to that expression.

$$\begin{aligned} r_{s,k} &= \frac{p(1-p) - \sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1}}{p(1-p)} \\ &= \frac{p - \sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1} - p^2}{p - \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} + \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - p^2} \\ &= \frac{\sum_{i=0}^n f_i \frac{i}{n} - \sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1} - p^2}{\sum_{i=0}^n f_i \frac{i}{n} - \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} + \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - p^2} \\ &= \frac{\sum_{i=0}^n f_i \frac{i}{n} \left(1 - \frac{n-i}{n-1}\right) - p^2}{\sum_{i=0}^n f_i \frac{i}{n} \left(1 - \frac{i-1}{n-1}\right) + \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - p^2} \\ &= \frac{\sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - p^2}{\sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1} + \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - p^2} \end{aligned}$$

Here $\sum_{i=0}^n f_i \frac{i}{n} \frac{n-i}{n-1}$ is the without replacement version of within-group variance as before. If we call $\sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - p^2 = \sum_{i=0}^n f_i \frac{i}{n} \frac{i-1}{n-1} - \left(\sum_{i=0}^n f_i \frac{i}{n}\right)^2$ the without replacement version of between-group variance, we would read this as

$$F_{ST} = \frac{\text{between-group variance}}{\text{average within-group variance} + \text{between-group variance}} \quad (49)$$

Again, for large group sizes n , the with or without replacement versions are close. The sampling in empirical studies typically seems to be without replacement anyway, so for all intents and purposes, these three definitions of relatedness are equivalent.

S.8.2 Empirical estimates

The condition for cooperation to be selected for by group selection used in, for instance, Bell et al. (2009) (see also Weir and Cockerham 1984; Crow and Aoki 1984; Aoki and Nozawa 1984; Bowles 2006; 2009; Langergraber et al. 2011; Walker 2014; Rusch 2018) is

$$\frac{\beta(w_g, p_g)}{\beta(w_{ig}, p_{ig})} > \frac{1 - F_{ST}}{F_{ST}}$$

Here $\beta(w_g, p_g)$ is the increase in the mean fitness of the group with an increase in the frequency of cooperators, or altruists, and $\beta(w_{ig}, p_{ig})$ is the fitness decrease of the individual acquiring the cooperative allele. The idea is that this criterion separates the fitness effects, on the left hand side of the inequality, from a measure that characterizes the population structure, on the right hand side of the inequality. In a setting with a linear public goods game, played within groups that compete with each other in a well mixed population of groups, such a separation can indeed be made in this way.

Suppose the fitness of a cooperator and a defector in a group with i cooperators (for cooperators: including the individual itself) are $w_{C,i} = 1 + \frac{i-1}{n-1}b - c$ and $w_{D,i} = 1 + \frac{i}{n-1}b$. Being a cooperator instead of a defector would then give $n - 1$ others in the group a fitness benefit of $\frac{1}{n-1}b$, adding up to an aggregate fitness benefit of b , at a fitness cost to the individual of c . The average fitness within the group would go from $\frac{iw_{C,i} + (n-i)w_{D,i}}{n} = 1 + \frac{i}{n}(b - c)$ to $\frac{(i+1)w_{C,i+1} + (n-i-1)w_{D,i+1}}{n} = 1 + \frac{i+1}{n}(b - c)$, which amounts to an increase of $\frac{1}{n}(b - c)$ as a result of an increase in the frequency of cooperators within that group of $\frac{1}{n}$. This makes $\beta(w_g, p_g) = b - c$. The fitness effect measured by $\beta(w_{ig}, p_{ig})$ is similarly interpreted as c . Rewriting $\frac{b-c}{c} > \frac{1-F_{ST}}{F_{ST}}$ as $(b - c)F_{ST} > c(1 - F_{ST})$ and then as $F_{ST}b > c$ gives Hamilton's rule, with $r = F_{ST}$.

If we were to measure $\beta(w_g, p_g)$ in a setting in which competition between groups is not actually global, but to some degree local, then the resulting value for $\beta(w_g, p_g)$ would not only reflect the effect of cooperators on the average fitness within the group, but a mixture of these fitness effects and the cancellation effect. A moderate value for $\beta(w_g, p_g)$ can both be the result of a moderate group benefit and the absence of the cancellation effect, and a high group benefit combined with the cancellation effect at the group level. In the latter case, the negative effect of having neighbouring groups with many cooperators, combined with the positive correlation between being a cooperative group and having neighbouring groups with many cooperators, would bias the estimated effect of – all else equal – the number of cooperators on average fitness within the group downwards. In other words, this term would end up absorbing the cancellation effect. In order to disentangle all fitness effects and the cancellation effect, one would have to estimate a more complex statistical model, which would not only use the composition of the own group as explanatory variable of the average fitness within the group, but also include the composition in neighbouring groups as an explanatory variable. This would be hard to estimate because it will require sufficiently high independent variation to overcome multicollinearity, but, if successful, it would separate the positive effect of the cooperation within the group from the negative effect of having a cooperative neighbouring group.

What most empirical papers do, however, is the estimation of only the F_{ST} , which is then taken as an indication of how conducive the population structure is to cooperation. We have seen that the absence or presence of the cancellation effect – which is part of the population structure – can however make a huge difference for how much the group needs to benefit from cooperators in it, relative to the individual costs, in order for cooperation to spread in the population.