A   Appendix

Many of the assumptions in section 3 were made for simplicity and ease of exposition. Here, we consider alternative assumptions and robustness of the model. Note that there are still many factors not included here, such as population structure and migration. These sorts of factors are not included so as to simplify the model and attempt to isolate conformist bias and relatedness as independent variables. One concern about the model is that migration should affect how some parts of the model work, notably how likely a trait is to be imitated based on conformism and the level of relatedness in a group. Below we vary the level of conformism, which also affects the likelihood a trait is imitated, and the level of relatedness in order to alleviate this concern.

First, we consider alternative methods of incorporating conformist bias. There are two questions to be interested in here. The first is: how likely it is that people pay attention to popularity vs. paying attention to payoffs? We considered this already in section 3, but here we look at a case where $\mathcal{C}_{optimal}$ increases at a higher rate, reaching (close to) its maximum at a lower group size:

$$\mathcal{C}_{optimal} = \frac{N}{N + 10} \tag{5}$$

We will call this the 'likelihood of conformism'.

The next question is: given that someone is paying attention to popularity rather than payoffs, how conformist will they be? Will they automatically adopt the most common trait, or will they just be more likely to adopt it the more common it is (and if so, how much more likely)? Equation (2) in section 3 was a very simple way of capturing this conformism, but as we noted it does not fit with the way many people talk about conformist bias, where traits that are common are imitated more often than their frequency in the population. This is better represented by having the likelihood of adopting a trait based on frequency be an s-shaped curve. To capture this, one can use an equation like the one found in Lehmann and Feldman (2008, p. 508):

$$c(x_a) = \frac{x_a^{\alpha}}{x_a^{\alpha} + (1 - x_a)^{\alpha}} \tag{6}$$

where, one might say that when $\alpha > 1$ there is conformist bias. By varying α we can vary how conformist people are when they are paying attention to popularity. See figure 8(a) the the shape this equation takes for various levels of α. We will call this the 'strength of conformism'.

We can then modify equation (2) so that the value of a trait includes both the likelihood and now the strength of conformism:

$$v_a(x(t)) = v_o + \mathcal{C} \cdot c(x_a(t)) + (1 - \mathcal{C}) \cdot u_a(x(t)) \tag{7}$$

Often, modelers will include some likelihood of individual versus social learning as well. These two possibilities are collapsed here into the second term of this
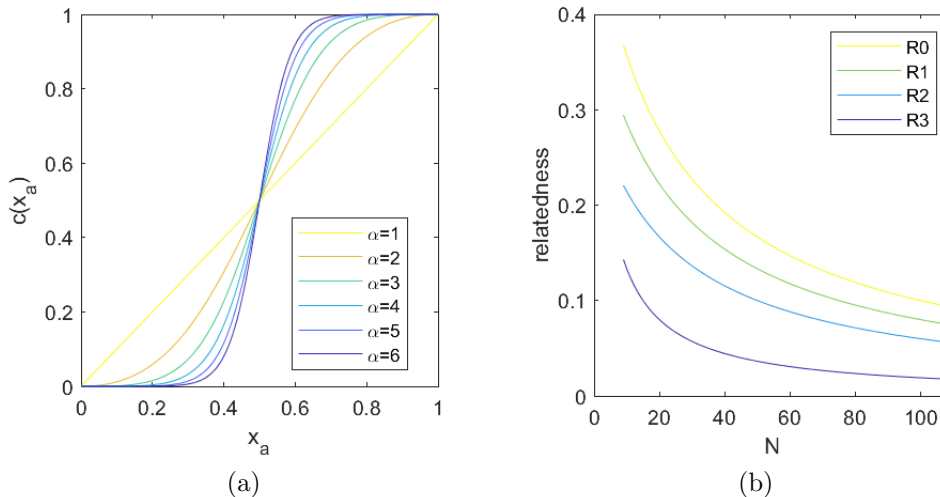
Figure 8: Values considered for (a) the strength of conformism and (b) the level of relatedness for various group sizes.

equation, which can both be interpreted as either individual learning based on payoffs or a form of non-conformist social learning like success-based imitation.

Here, we also vary the form of the relatedness equation. As mentioned in section 3, the actual level of relatedness depends on many factors, including migration, endogamy, dispersal, etc. Here we keep the same general form – relatedness starts fairly high in small groups but decreases as group size increases – but vary the initial level of relatedness and the speed at which it decreases. See figure 8(b) for the options considered for how relatedness changes as group size changes.[22]

We are interested in how long phenotypic altruism can persist, given these differing assumptions about conformist bias and relatedness. For that reason, we let the simulations run longer (3,000 generations), though we kept the maximum group size at 109 as in section 3.[23] If the altruistic phenotype ever dropped below 50% of the population, altruism was considered to be longer maintained at that point in time based on the reasoning that it would not be able to be sustained by conformist bias in the future. Figure 9 shows these results.

We can see that, for all the considered options for tracking relatedness over time, how long altruism can be maintained depends on how quickly conformist bias evolves, as in section 3. The faster it evolves, the longer altruism is maintained. We can additionally see that altruism can be maintained longer the more conformist people are when they do pay attention to popularity, again, for

---

[22]The equations generating these lines are: $R0 = \frac{.5}{1+.4N}, R1 = \frac{.4}{1+.4N}, R2 = \frac{.3}{1+.4N}$, and $R3 = \frac{.4}{1+.2N}$.

[23]Letting the group size continue to increase does not significantly affect the results, though.
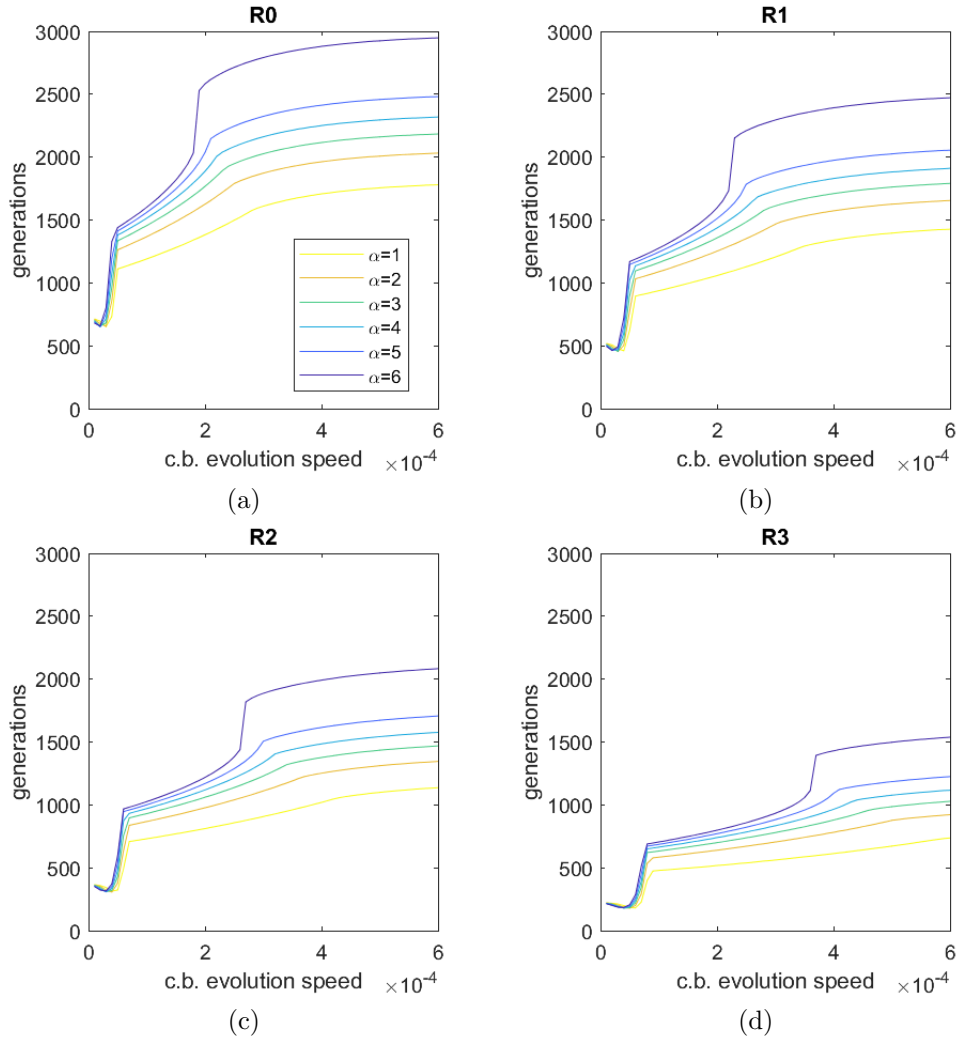
Figure 9: Length of time phenotypic altruism is maintained under various assumptions regarding relatedness and conformist bias (the speed it evolves toward the optimal $\mathcal{C}$ and the strength of conformism, α), for $c = .1$, $b = .4$, and $v_0 = f_0 = 1$.

each of R0-R3. So, the qualitative results we described in section 3 are robust to differing assumptions regarding the level of relatedness and the strength of conformism, α. Of course, the higher relatedness is, the longer altruism can last as well, as figure 9(a) shows results where there are higher levels of relatedness for longer than in 9(b), and so on compared to 9(c) and 9(d) as well.

There are two features of these results that deserve mention. First, the length of time altruism can be maintained dips for small speeds of the evolution of $\mathcal{C}$. That is, somewhat unintuitively, altruism can actually last a bit longer when there is virtually no conformism. This is because without any conformist bias, genetic altruists switch most quickly to phenotypic non-altruism and therefore hardly ever pay the costs associated with the altruistic phenotype. So, fitness differences between the genes are lower, and genetic altruism (and, as a result, phenotypic altruism) can be maintained for slightly longer periods of time. This is somewhat similar to the phenomenon we noted in section 3, where the altruistic genotype can be maintained longer under slow conformist bias evolution than for intermediate conformist bias evolution speed.

Second, in many cases, there seems to be a point at which speeding up conformist bias evolution just a little increases the length of time altruism is maintained by a lot. This point is different for different relatedness equations and values of α. Comparing the fast to the intermediate case in section 3 can give us a hint at why this might occur. In the fast case, the value of altruism rose above non-altruism before relatedness dropped too low to sustain genetic altruism. In the intermediate case, the value of altruism overtook the value of non-altruism after genetic altruism was no longer favored. So while altruism was maintained for a period, there was an extra advantage in the fast case because conformist bias was high enough at a point where there were still many genetic altruists introduced into the population every generation. It is possible that a similar effect explains the extra boost altruism gets from some slight increases in the speed of conformist bias evolution – the increase in speed allows the value of altruism to overtake the value of non-altruism before genetic altruism is no longer favored.

Finally, one might wonder whether there are any cases where altruism can be sustained indefinitely. This can happen if relatedness is high enough for long enough and high levels of conformist bias evolve quickly. For instance, when relatedness is described by equation R0, $\alpha = 7$, and the speed of conformist bias evolution is $\geq .0018$, altruism is maintained for the entire length of a 1,000,000 generation simulation. Other parameter combinations can generate this long-lasting altruism as well: as long as altruism is sufficiently frequent at the time when a strong enough conformist bias arises, altruism can be maintained indefinitely.