2023

A kin-selection model of fairness in heterogeneous populations

Geoff Wild
Western University, gwild@uwo.ca

Veronica J. Flear

Graham J. Thompson

A kin-selection model of fairness in heterogeneous populations

Geoff Wild [a,*], Vonica J. Flear [b], Graham J. Thompson [b]

[a] *Department of Mathematics, Western University, 1151 Richmond Street, London, N6A 5B7, Ontario, Canada*
[b] *Department of Biology, Western University, 1151 Richmond Street, London, N6A 5B7, Ontario, Canada*

ARTICLE INFO

ABSTRACT

Humans and other primates exhibit pro-social preferences for fairness. These preferences are thought to be reinforced by strong reciprocity, a policy that rewards fair actors and punishes unfair ones. Theories of fairness based on strong reciprocity have been criticized for overlooking the importance of individual differences in socially heterogeneous populations. Here, we explore the evolution of fairness in a heterogeneous population. We analyse the Ultimatum Game in cases where players' roles in the game are determined by their status. Importantly, our model allows for non-random pairing of players, and so we also explore the role played by kin selection in shaping fairness. Our kin-selection model shows that, when individuals condition their behaviour on their role in the game, fairness can be understood as either altruistic or spiteful. Altruistic fairness directs resources from less valuable members of a genetic lineage to more valuable members of the same lineage, whereas spiteful fairness keeps resources away from the competitors of the actor's high-value relatives. When individuals express fairness unconditionally it can be understood as altruistic or selfish. When it is altruistic, unconditional fairness again serves to direct resources to high-value members of genetic lineages. When it is selfish, unconditional fairness simply improves an individual's own standing. Overall, we expand kin-selection based explanations for fairness to include motivations other than spite. We show, therefore, that one need not invoke strong reciprocity to explain the advantage of fairness in heterogeneous populations.

1. Introduction

Humans and certain non-human primates show preferences for fairness even when fairness seems to be at odds with personal success (Fehr and Fischbacher, 2003; Proctor et al., 2013). Such preferences are thought to be a reflection of an underlying prosocial tendency to engage in strong reciprocity (Gintis, 2000). Strong reciprocators reward cooperative neighbours and punish non-cooperative ones, so strong reciprocity can be viewed as an individual-level adaptation that promotes cooperative social norms.

Some of the most conspicuous examples of strong reciprocity come from experiments involving the Ultimatum Game (Güth et al., 1982). The game is played by two individuals who must split a resource of fixed value. One member of the pair (the *proposer*) offers some fraction of the resource to its partner and earmarks the remaining fraction for itself. If the partner (the *responder*) accepts the proposer's offer, then the resource is divided between the individuals accordingly. However, if the responder refuses, then both individuals receive nothing. Importantly, both proposer and responder understand, in advance, all consequences of accepting or refusing a proposal.

Given the set-up of the Ultimatum Game, it is clear that a rational responder ought to accept any non-zero fraction of the resource offered by the proposer, and that a rational proposer ought to offer the smallest fraction possible to the responder. Nevertheless, human responders rarely accept offers that see them receive less than one-quarter of the resource, and human proposers most often offer an even split (Fehr and Fischbacher, 2003). Similar patterns have been reported for chimpanzees, *Pan troglodytes* (Proctor et al., 2013).

The general way in which both humans and non-human primates approach the Ultimatum Game is certainly curious, but so too is the variation evident in their respective approaches. For example, male humans tend to be more generous toward partners that possess certain qualities like attractiveness and higher social status (Eisenbruch et al., 2016). In addition, it has been argued that payoffs for chimpanzees in the Ultimatum Game can be influenced by dominance rank (Proctor et al., 2013). The argument for chimpanzees, here, is strengthened by aggressive acts perpetrated by responders (Proctor et al., 2013), combined with the fact that aggression is intimately connected to dominance in primates (de Waal, 1986).

Despite the importance of individual quality to the Ultimatum Game and our understanding of how it is played, theoretical exploration in heterogeneous populations is still lacking. Previous modelling work has dealt with heterogeneity of individuals playing this game

---

by considering average effects (Page and Nowak, 2000). Other work has explicitly recognized social heterogeneity, in particular, but has also shown that additional spatial structure is needed to incentivize fairness (Killingback and Studer, 2001). To our knowledge, though, theoretical treatments have yet to explore heterogeneity in the Ultimatum Game in a way that allows one to easily outline the adaptive significance of fairness, i.e. the purpose thereof (sensu Gardner, 2017). This leaves open the question, what goal does fairness achieve in heterogeneous populations?

In this paper we investigate the adaptive significance of fairness in the Ultimatum Game with a simple mathematical model that pits a fair strategy against a rational one. We construct an expression for the co-variance between fitness and fairness, as has been done elsewhere (e.g. Gintis, 2000). We then use that expression to derive a statement about kin selection. Like previous authors (Lehmann et al., 2007; West et al., 2008), we find that kin selection can be used in place of strong reciprocity to characterize the advantage of fairness. We also show that, in a heterogeneous population, kin-selection based explanations for the advantage of fairness can be described as altruistic, spiteful, or selfish. In particular, altruistic fairness in a heterogeneous population disproportionately rewards high-value members of a genetic lineage, whereas spiteful fairness disproportionately punishes competitors of high-value members of a genetic lineage.

2. Model

We consider a population of haploid asexual individuals who can be placed into one of two categories, e.g. high status or low status, dominant or subordinate, large or small, etc. Admittedly, haploid asexual genetics does not reflect the biology of humans and other primates. Such a system of inheritance, though, makes our model simple without compromising our ability to comment on taxonomic groups of interest. We defer discussion of this point to the final section of the paper.

Individuals form pairs to decide how to divide a resource of value h. To make their decision, each partnership plays the Ultimatum Game (Güth et al., 1982). For us, though, partners always belong to different categories. Furthermore, the role each partner adopts in the Ultimatum Game is determined by the category to which it belongs. For example, proposers might always be low-status individuals and responders might always be high-status individuals. It follows that we can think of one category of individual as being synonymous with the proposer role, and the other category of individual as being synonymous with the responder role. Similar links between status and role have been made in a previous study of the Ultimatum Game (Killingback and Studer, 2001).

We assume that an individual's approach to the Ultimatum Game is determined by its genotype at a single, diallelic locus. Individuals carrying the a allele at this locus approach the game rationally. In other words, they (i) offer to give their partner a fraction $\varepsilon < \frac{1}{2}$ of the resource when in the proposer role and (ii) accept any offer that leaves them with some non-zero fraction of the resource when in the role of responder. By contrast, individuals carrying the A allele take a fair approach, in that they (i) offer to split the resource evenly when in the proposer role and (ii) only accept offers that leave them with at least half of the resource when acting as responder. In some cases, we allow individuals to express their alleles only when they adopt a particular role. In those cases of conditional expression, individuals who adopt the other role are carriers of the alleles only.

Let Δp_{pr} and Δp_{res} denote the change in the frequency of the fair A allele that occurs among proposers and responders, respectively, over the course of a single generation. Following previous work (Taylor, 1990; Grafen, 2018), we focus on a weighted allele-frequency change, $\Delta p = \pi \Delta p_{\mathrm{pr}} + (1 - \pi) \Delta p_{\mathrm{res}}$. The weight π lies between 0 and 1. It measures the total reproductive value of individuals in the proposer role, defined as the probability that an allele chosen from a population of descendants far in the future will have originated from the current

pool of proposers. Similarly, the weight $1 - \pi$ measures the total reproductive value of responders. We use the expression $1 - \pi$ to reflect the fact that a descendant allele that did not originate from a present-day proposer must have originated from a present-day responder. Overall, the weights π and $1 - \pi$ allow our model to capture the idea that a change in the frequency of A among proposers may not have the same long-term evolutionary implications as an identical change in the frequency of A among responders. Because proposers and responders occur with equal frequency in this model (one is always paired with the other) we can also understand π and $1 - \pi$ as measuring the reproductive value of one proposer and one responder, respectively. If $\pi = 1 - \pi = \frac{1}{2}$, then, the proposer and responder have the same evolutionary significance and we have no real heterogeneity. As this is outside the scope of the paper we explore the case $\pi = \frac{1}{2}$ in Appendix A.

We can derive a simple expression for Δp based on the Price (1970) equation (see details in Appendix A), but first we need some additional notation. We use W_{pr} to represent the fitness of a given proposer, measured relative to the average proposer. Similarly, we use W_{res} to represent the fitness of a given responder, measured relative to the average responder. We define the fitness of a given individual, regardless of role, as $W = \pi W_{\mathrm{pr}} + (1 - \pi) W_{\mathrm{res}}$. Finally, we use G to represent the genotype of a given individual, with $G = 0$ for those carrying the a allele, and $G = 1$ for those with A. Using our new notation we obtain

$$\bar{W} \Delta p = \mathrm{Cov}(W, G) \tag{1}$$

where $\bar{W}$ is expected value of W. Eq. (1) predicts that the frequency of A increases if and only if the presence of this allele is positively correlated with fitness. Importantly, Eq. (1) neglects mutation, and so the weighted allele-frequency change that appears in this line is more accurately interpreted as a partial change due to selection.

3. Inclusive-fitness analysis

Computing the covariance in Eq. (1) is difficult, in general, but is made easier if we assume selection is weak. Because weak selection can arise in different ways (Wild and Traulsen, 2007), we stress that the way we use it here assumes that h, the value of the resource to be split, contributes only a small amount to an individual's fitness. Under this version of weak selection, and following results presented elsewhere (Taylor, 1990, 2017; Gardner et al., 2011; Grafen, 2018), it can be shown that

$$\mathrm{Cov}(W, G) \propto \left(-C + \frac{D}{2} \right) + \left(B + \frac{D}{2} \right) R \tag{2}$$

(see Appendices). Here, C represents the additive fitness cost of switching from the rational strategy (allele a) to the fair strategy (allele A), when partnered with an individual using the rational strategy. Along similar lines, B represents the additive fitness benefit of switching from rational to fair, when partnered with a rational individual. Where possible, D represents the non-additive synergistic changes in fitness achieved by rational pairs of individuals who jointly switch to fair strategies. The coefficient R gives the relatedness between pairs. Note that $R = 0$ means that partners are no more similar to each other than they are to the average individual in the population (Grafen, 1985b). With this in mind, R approaches -1 when partners are very dissimilar (non-kin) and approaches 1 when they are very similar (close kin).

The right-hand side of line (2) looks like a version of Hamilton's (1964) rule, with cost and benefit terms modified by each partner's share of the synergy they generate. The similarity to Hamilton's rule is, of course, not accidental as we can accurately describe Eq. (2) as the inclusive-fitness effect of fairness. In fact, we might re-write (2) as $-C + BR + D(1 + R)/2$ to make it obvious that D can be interpreted as a benefit awarded to the average member of a pair. Readers familiar with work by Queller (e.g. Queller, 1985, 2011) will note that no coefficient of synergy appears in (2). For those readers we emphasize that the coefficient of synergy has not been omitted, rather the weak
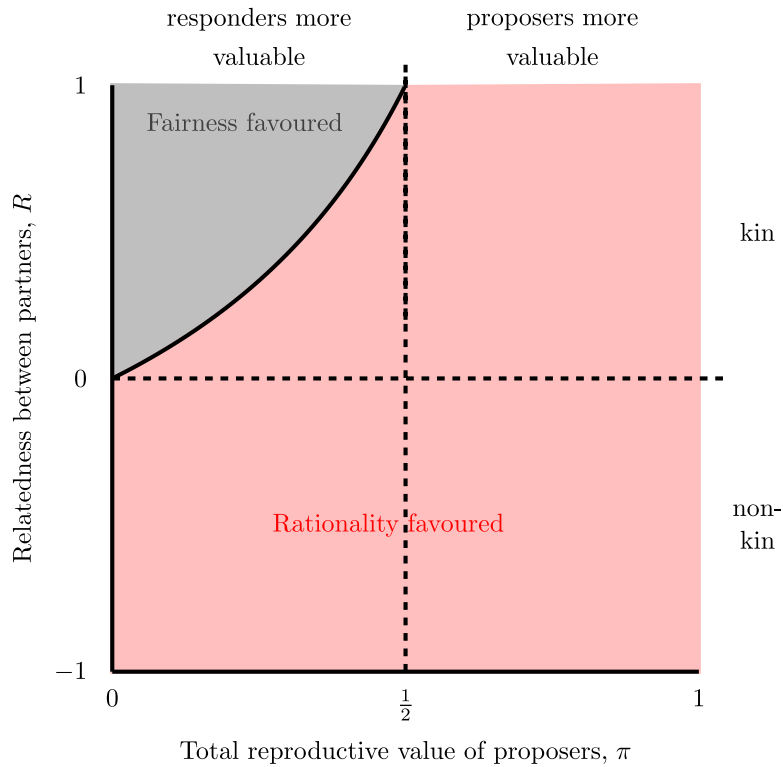
Fig. 1. Summary of model predictions when only the proposer expresses the fair allele *A*. Fairness in the Ultimatum Game is favoured over rationality in the grey region. The opposite is true in red regions. In this scenario, fairness can be understood as a tendency toward altruism. Where it is favoured, then, fairness directs resources away from the low-value proposer and toward its high-value kin.

selection assumption simply allows it to be subsumed by R, as shown in Appendix C.

For the Ultimatum Game played among individuals who may be either rational (*a* carriers) or fair (*A* carriers) we find that if alleles are expressed only by proposers, then

$$
\left.
\begin{array}{rcl}
C &=& \pi\left((1-\varepsilon)-\frac{1}{2}\right) \\
B &=& (1-\pi)\left(\frac{1}{2}-\varepsilon\right) \\
D &=& 0.
\end{array}
\right\}
\tag{3}
$$

Here, the constant C specifically captures the additive reduction in reproductive value experienced by a focal proposer who has decided to switch from a rational strategy to a fair one. The constant B represents the additive increase in reproductive value experienced by the responder paired with the focal proposer. In this case, there is no scope for non-additive synergistic change, as the responder does not express the fair A allele. After substituting (3) in (2), we find that fairness, expressed only by proposers, is favoured by selection whenever

$$
\left(\frac{1}{2}-\varepsilon\right)\left(-\pi+(1-\pi)R\right)>0.
\tag{4}
$$

Of course, fairness is disfavoured when the preceding inequality is reversed.

Now, if alleles are expressed only by responders, we obtain

$$
\left.
\begin{array}{rcl}
C &=& (1-\pi)\varepsilon \\
B &=& -\pi(1-\varepsilon) \\
D &=& 0.
\end{array}
\right\}
\tag{5}
$$

When the previous expressions are substituted into (2), we see that fairness, expressed only by responders, is favoured by selection whenever

$$
-(1-\pi)\varepsilon-\pi(1-\varepsilon)R>0
\tag{6}
$$

and disfavoured when the inequality is reversed.

Finally, when fairness is expressed by both proposer and responder we find

$$
\left.
\begin{array}{rcl}
C &=& \pi\left((1-\varepsilon)-\frac{1}{2}\right)+(1-\pi)\varepsilon, \\
B &=& (1-\pi)\left(\frac{1}{2}-\varepsilon\right)-\pi(1-\varepsilon), \\
D &=& \pi(1-\varepsilon)+(1-\pi)\varepsilon.
\end{array}
\right\}
\tag{7}
$$

Note that to express the true non-additive consequences, represented by D, we must subtract the additive fitness effects of switching from the total fitness change owing to a joint move away from rationality and toward fairness. Substituting (7) into (2) leads to

$$
-\varepsilon\left(\frac{1}{2}-\pi\right)+(1-\varepsilon)\left(\frac{1}{2}-\pi\right)R=\left(\frac{1}{2}-\pi\right)\left(R-\frac{\varepsilon}{1-\varepsilon}\right)(1-\varepsilon)>0
\tag{8}
$$

as the condition for fairness to be favoured.

4. Results

4.1. Either proposer or responder express fairness, not both

When only the proposer expresses the fairness allele *A*, condition (4) shows fairness is advantageous when partners are kin and when the responder has high reproductive value (Fig. 1). In this case we assert that the purpose of fairness is to altruistically direct resources from a low-value actor (the proposer) to a genetic relative of high value (the responder).

When only the responder expresses fairness, condition (5) shows fairness is advantageous only when partners are non-kin and when the proposer has sufficiently high reproductive value (Fig. 2). Here, we assert that the purpose of fairness is to spitefully rob resources from a non-relative of sufficiently high-value (the proposer), by rejecting unfair proposals. Importantly, what we mean by 'sufficiently high value' depends on ε: as $\varepsilon \to \frac{1}{2}^-$ fairness is only favoured when non-kin proposers are more valuable; as $\varepsilon \to 0^+$ fairness is favoured for interactions with non-kin proposers of any possible value. Overall,
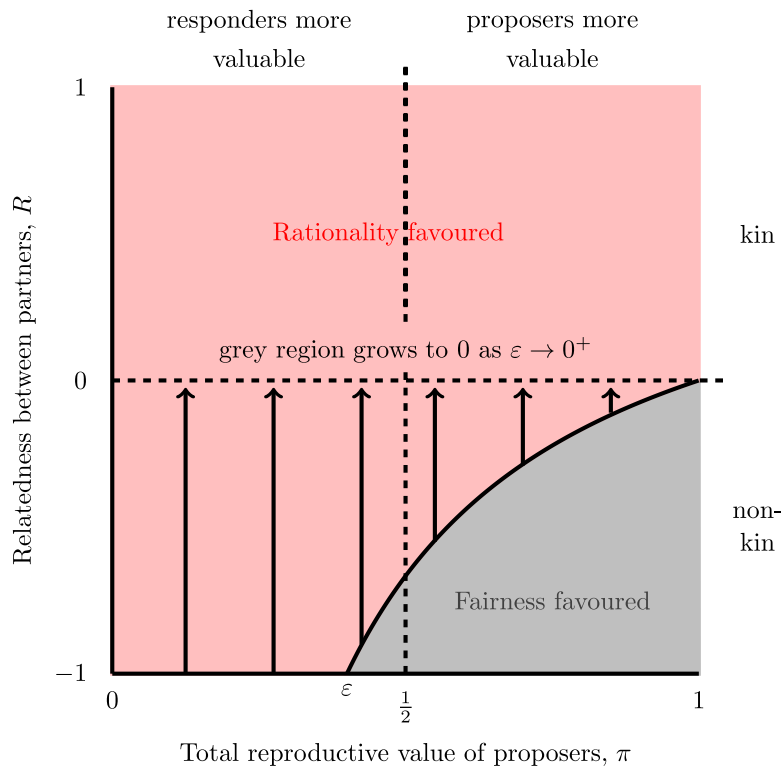
Fig. 2. Summary of model predictions when only the responder expresses the fair allele A. Fairness in the Ultimatum Game is favoured over rationality in the grey region. The opposite is true in red regions. In this scenario, fairness can be understood as a tendency toward spite. Where it is favoured, then, fairness robs resources from high-value non-kin. Arrows show that the grey region expands to fill the lower half of the sketch as $\varepsilon \to 0^+$.

the spiteful act of rejecting an unfair proposal indirectly improves the competitive prospects of the actor's relatives who currently find themselves in the proposer role.

4.2. Both proposer and responder express fairness

Matters are more complicated when both proposer and responder can express fairness. Condition (8) shows that fairness, in this case, is favoured when $\frac{1}{2} - \pi$ and $R - \frac{\varepsilon}{1-\varepsilon}$ have the same sign. In other words, fairness is favoured when either (i) the proposer is less valuable than the responder ($\frac{1}{2} > \pi$) and partners are close kin ($R > \frac{\varepsilon}{1-\varepsilon}$), or (ii) the proposer is more valuable than the responder ($\frac{1}{2} < \pi$) and partners are distant or non-kin ($R < \frac{\varepsilon}{1-\varepsilon}$) (Fig. 3).

The synergy present in the model suggests the inclusive-fitness narrative surrounding the advantage of fairness might be awkward, or even unattainable. Strictly speaking, the inclusive-fitness effect of a behaviour should ignore fitness effects that are received from the focal actor's social environment (Hamilton, 1964). By its very nature, though, synergy captured by the D term requires us to consider simultaneous deviant action from both actor and its partner. We get around this problem by modifying costs to include frequency-dependent effects necessary to deal with synergy (Gardner et al., 2011). In other words, we simply use $-C + \frac{D}{2}$ and $B + \frac{D}{2}$ as the cost and benefit of fairness, respectively, and put these into Hamilton's rule.

From line (8) we find that the cost term

$$-C + \frac{D}{2} = -\varepsilon \left(\frac{1}{2} - \pi \right) \tag{9}$$

is negative only when the proposer is less valuable than the responder; when the opposite is true, this so-called cost is actually a benefit reaped by the actor. Similarly, the benefit term in (8),

$$B + \frac{D}{2} = (1 - \varepsilon) \left(\frac{1}{2} - \pi \right), \tag{10}$$

is only a true benefit when the proposer is less valuable than the responder. When the proposer is more valuable than the responder, the so-called benefit is really a cost.

With costs and benefits laid out, we argue that the advantage of fairness when the proposer is less valuable can be understood as an altruistic act: the average fair individual tends to pay a cost to provide a benefit to a close relative. Referencing the top left quadrant of Fig. 3, we go further and say that, again, fairness serves the purpose of directing resources from less valuable individuals to the more valuable members of the same lineage.

When the proposer is more valuable, fairness achieves an advantage because it generates a benefit (i.e., a false cost) for the average fair individual while simultaneously imposing a cost (i.e., a false benefit) on distant-kin and non-kin competitors. Evidently, fairness can simply be characterized as selfish in this scenario.

5. Discussion

Conspecifics are rarely identical. In some cases the differences between them not only reflect social structure, but also result in divergent abilities to acquire and hold resources, or share resources with their neighbours (Sapolsky, 2005; Watts, 2010; Sánchez-Amaro et al., 2018; Zeng et al., 2022). Such differences occur even among laboratory-housed primates similar to those used in studies of pro-social behaviour (Boccia et al., 1988). Given that heterogeneity is pervasive and can be linked to social interactions, we wanted to know how it affects the evolution of pro-social preferences for fairness. We asked specifically, what adaptive purpose does fairness serve in a heterogeneous population?

To answer our question we analysed the well-known Ultimatum Game in the case where individuals who adopt a given role are not the evolutionary equivalent of those who adopt the opposing role. For example, socially dominant individuals might be proposers disproportionately often because they are better able to acquire resources,
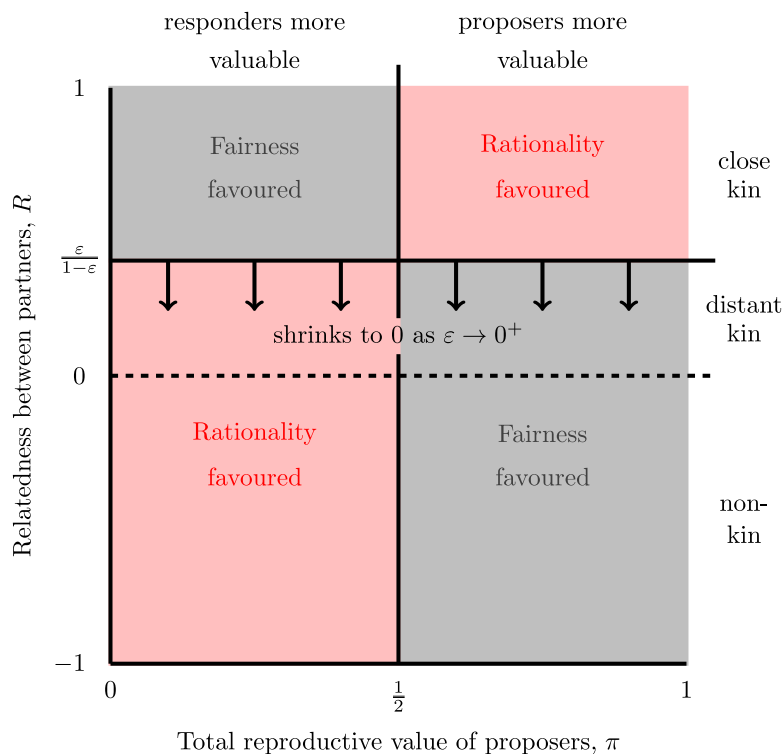
Fig. 3. Summary of model predictions when both proposers and responders can express the fairness allele *A*. Fairness in the Ultimatum Game is favoured over rationality in grey regions. The opposite is true in red regions.

or to oversee how resources are distributed. Alternatively, physically dominant individuals might be responders disproportionately often because they can acquire resources indirectly from subordinates with relatively little effort.

Previous authors have interpreted fairness in the Ultimatum Game as a tool for reputation management (Milinski, 2013). Under this view, fairness is enforced by concerns about future punishment from strong reciprocators, but the view itself is not universally accepted. Opponents claim that explanations based on strong reciprocity create confusion that can only be resolved by framing the advantage of fairness using 'kin selection' terms, most notably 'spite' (Lehmann et al., 2007; West et al., 2008). Others claim that strong reciprocity ignores the importance that players place on maintaining their standing in socially heterogeneous populations (Yamagishi et al., 2012). Our results have combined elements from both objections and framed the purpose of fairness in heterogeneous populations using kin-selection language.

For heterogeneous populations we find that fairness, when it is expressed only by individuals in one particular role, serves to either direct resources toward the actor's more valuable relatives, or keep resources from the competitors of more valuable relatives. In the former instance, fairness can be described as an act of altruism, and in the latter instance it can be described as an act of spite. Our use of typical 'kin selection' terms like 'altruism' and 'spite' is facilitated by the fact that conditional expression of fairness in this model eliminates the synergistic fitness effects that can sometimes confound Hamilton's rule (Queller, 1985; but see Grafen, 1985a and Gardner et al., 2011). This same observation has been made in other settings where synergy plays a noticeable role (Queller, 1996), but here we take it as evidence that fairness need not be a reflection of strong reciprocity. In this way, we align with points made in Lehmann et al. (2007) and West et al. (2008), but also we reveal that kin-selection forces can allow fairness to be viewed as altruistic, not just spiteful.

When fairness can be expressed by individuals in both roles, synergy between partners can be realized. For this case, we provided a version of Hamilton's rule that divides the synergistic fitness effects equally between a fair actor and its partner (it was, in fact, a simple generalization of a condition provided by Gardner et al. 2011 and Taylor 2017). Our version of the rule exploited the weak-selection assumption to effectively hide the allele-frequency dependence inherent to the model, which meant we could state simple conditions for the advantage of fairness. Those conditions again show that fairness can be an altruistic act whose purpose is to see that resources are directed to higher-value members of the same genetic lineage. We also find that fairness can be understood as a selfish act under some conditions, but in all cases where fairness can be expressed in both roles its advantage can be described using kin-selection terminology only.

Overall, our findings outline the adaptive purpose of fairness in heterogeneous populations. They point to the reasons why factors like spatial structure – factors that affect patterns of genetic relatedness in populations – have been seen to promote the evolution of fairness (Killingback and Studer, 2001). By uncovering the logic behind fairness, we have helped future authors predict the fate of fairness in new situations where specific theoretical exploration is lacking.

It is interesting to ask, does there exist a cultural evolutionary process that might mirror the process we explore here? The simple answer is, yes. The mathematical formalism upon which our model rests, namely the Price (1970) equation, is not restricted to genetic systems, as Price himself acknowledged in his original work. When framing the problem as cultural evolutionary process, we would shift focus away from a genetic locus, where alleles are found, to a cultural locus, where ideas are found (Lehmann et al., 2008). Relatedness, *R*, would then be a measure the extent to which partners' strategies are culturally correlated. The interpretation of the coefficient *π* would also change: in a cultural model *π* would represent the cultural influence of the proposer.

Our model assumed individuals are haploid and asexual, and this is certainly at odds with the biology of humans and other primates. Nevertheless, our simple assumptions about individuals' genetics do not compromise our ability to apply our findings broadly. In evolutionary game theory, equilibrium predictions made by haploid asexual models

match those made by one-locus diploid models (Geritz and Kisdi, 2000). And while coefficients of relatedness definitely change with the genetic system one considers (e.g. Taylor, 1988), the fact that we treat these coefficients as model inputs means the model itself is sufficiently flexible to handle whatever value one may care to study (haploid or diploid) (Cooper et al., 2018). In general, though, we recognize that we are – as Grafen (1984) points out – taking an often-justifiable "leap of faith" that the genetic system underlying fairness will find the same strategies as those found by the asexual agents envisioned by our model.

As mentioned, our analysis also relied on a weak-selection assumption. That assumption allowed us to recover a version of Hamilton's rule that incorporated synergy, but we did not make the assumption to recover Hamilton's rule. In fact, the weak-selection assumption allowed us to express fitness differences among individuals as simple functions of payoffs arising from the Ultimatum Game. Simply put, our version of Hamilton's rule was a by-product of an assumption we needed for other reasons. It could be that a notion of strong reciprocity – one that is distinct from altruism or spite – is required when selection is strong. Nevertheless, it seems that strong reciprocity can be ignored in the weak-selection domain. Previous theoretical and empirical work has demonstrated that the size of the resource to be split can influence how the Ultimatum Game is played (Härdling, 2007; Andersen et al., 2011), especially when the resource is large enough to impact individual success significantly (Cameron, 1999). It would, therefore, be interesting to explore adaptive explanations of fairness based on strong reciprocity and under strong selection. Those explorations, however, will have to contend with the challenge of clearly linking individual fitness to payoffs associated with specific social interactions if something other than random interactions is studied. What is more, it is unclear how important windfalls, like those presented to players in experimental versions of the Ultimatum Game (e.g. Cameron, 1999), have been in shaping the evolution of pro-social tendencies. So, while effects of resource size reported by others cannot be denied, we might speculate that they are artificial and have little to do with the adaptive significance of fairness established over the course of species' evolution.

CRediT authorship contribution statement

Geoff Wild: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition. **Vonica J. Flear:** Methodology, Formal analysis, Writing – review & editing. **Graham J. Thompson:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Fitness and allele-frequency change

Individuals receive a payoff from playing the Ultimatum Game, and this payoff contributes to its fitness. We use w_j to denote the payoff expected by a focal individual in role $j = $ pr, res, and we use $\bar{w}_j$ to denote the payoff expected by the average individual in role j. We then express fitness of the focal individual in role j as

$$W_j = \frac{1 + \delta w_j}{1 + \delta \bar{w}_j} \tag{A.1}$$

where the constant $\delta > 0$ describes the extent to which payoff affects fitness. If δ is small, we say selection is weak because the Ultimatum Game contributes little to fitness. We stress that this is not weak selection in the usual sense: allele A does not have low phenotypic penetrance (Wild and Traulsen, 2007).

With our version of the weak-selection assumption in place, we approximate fitness as

$$W_j \approx 1 + \delta(w_j - \bar{w}_j). \tag{A.2}$$

We now introduce $w_{j,X|Y}$ to denote the payoff obtained by a focal individual in role j who carries allele $X = a, A$ when partnered with an individual in the opposite role, $\neg j$, who carries allele $Y = a, A$. We can use this new piece of notation to express the fitness residual of a focal individual as

$$\delta(w_j - \bar{w}_j) = \delta \underbrace{(w_{j,A|a} - w_{j,a|a})}_{-C_j} G_j + \delta \underbrace{(w_{j,a|A} - w_{j,a|a})}_{B_{\neg j \to j}} G_{\neg j}$$
$$+ \delta \Bigg(\underbrace{(w_{j,A|A} - w_{j,a|a}) - (w_{j,A|a} - w_{j,a|a}) - (w_{j,a|A} - w_{j,a|a})}_{D_j} \Bigg) G_j G_{\neg j} \tag{A.3}$$

where G_j and $G_{\neg j}$ are random variables that indicate the presence of the fair A allele in the focal individual and its parnter, respectively. Note that the fitness deviation written in this new form emphasizes neighbour-modulated effects on the payoff expected by the focal individual. It includes additive effects of the individual's own actions (term labelled C_j), additive effects owing to the partner's actions (term labelled $B_{\neg j \to j}$), and the non-additive consequences of joint action of both partners (term labelled D_j). It is instructive to note that the terms $-C_j$, $B_{\neg j \to j}$, and D_j can be understood as coefficients from a least-squares regression of response variable $w_j - \bar{w}_j$ on predictors G_j, $G_{\neg j}$, and $G_j G_{\neg j}$.

If we now calculate $\mathrm{Cov}(W, G \,|\, \text{focal ind. in role} = j) \overset{\mathrm{def}}{=} \mathrm{Cov}(W_j, G_j)$ we find

$$\mathrm{Cov}(W_j, G_j)$$
$$= \delta \Bigg(-C_j \underbrace{\mathrm{Cov}(G_j, G_j)}_{= \mathrm{Var}\, G_j} + B_{\neg j \to j} \mathrm{Cov}(G_j, G_{\neg j}) + D_j \mathrm{Cov}(G_j, G_j G_{\neg j}) \Bigg)$$
$$= \delta \, \mathrm{Var}\, G_j \left(-C_j + B_{\neg j \to j} \frac{\mathrm{Cov}(G_j, G_{\neg j})}{\mathrm{Var}\, G_j} + D_j \frac{\mathrm{Cov}(G_j, G_j G_{\neg j})}{\mathrm{Var}\, G_j} \right). \tag{A.4}$$

We recognize $R_j = \frac{\mathrm{Cov}(G_j, G_{\neg j})}{\mathrm{Var}\, G_j}$ as the coefficient of relatedness between partners (Michod and Hamiton, 1980), but from the perspective of the individual in role j. We also recognize $S_j = \frac{\mathrm{Cov}(G_j, G_j G_{\neg j})}{\mathrm{Var}\, G_j}$ as the coefficient of synergy between partners (Queller, 1985), but again from the perspective of the individual in role j. If p_j is the expectation of G_j and f the expectation of $G_j G_{\neg j}$, then we can also write

$$R_j = \frac{f - p_j p_{\neg j}}{p_j(1 - p_j)} \qquad \text{and} \qquad S_j = \frac{f}{p_j} \tag{A.5}$$

as alternative forms.

We also recognize that

$$\mathrm{Cov}(E\{W|\text{focal ind. in role } j\}, E\{G|\text{focal ind. in role } j\}) = 0 \tag{A.6}$$

owing to the fact that $E\{W|\text{focal ind. in role } j\} = 1$. It follows from a standard partition of covariance that

$$\mathrm{Cov}(W, G) = E\mathrm{Cov}(W, G \,|\, \text{focal ind. in role} = j)$$
$$= \pi_{\mathrm{pr}} \mathrm{Cov}(W_{\mathrm{pr}}, G_{\mathrm{pr}}) + \pi_{\mathrm{res}} \mathrm{Cov}(W_{\mathrm{res}}, G_{\mathrm{res}}) \tag{A.7}$$

where π_j is the probability mass we assign to the covariance in role j. The probability mass used in (A.7) must be the limiting distribution of the Markov chain with transition matrix

$$T = \begin{bmatrix} p_{\text{pr}\leftarrow\text{pr}} & 1 - p_{\text{pr}\leftarrow\text{pr}} \\ 1 - p_{\text{res}\leftarrow\text{res}} & p_{\text{res}\leftarrow\text{res}} \end{bmatrix} \tag{A.8}$$

where $p_{j\leftarrow j}$ gives the probability that an allele carried by an individual in role j descended from one carried by an individual in role j in the previous generation (Taylor, 1990).

In biological terms, the elements of the limiting distribution, π_{pr} and π_{res}, would give the total reproductive value of individuals in the different roles if T remained fixed from one generation to the next. In reality, though, the entries of T can change over time; nevertheless, its limiting distribution is still helpful. Specifically, the distribution allows us to weigh the allele-frequency changes we have most recently observed – namely Δp_{pr} and Δp_{res} – to determine whether they combine in a way that generates a selective advantage for A (Grafen, 2015). This is the motivation behind defining the weighted allele frequency, $\Delta p = \pi_{\text{pr}}\Delta p_{\text{pr}} + \pi_{\text{res}}\Delta p_{\text{res}}$. The same definition also permits us to say

$$\Delta p = \text{Cov}(W, G) = \delta \pi_{\text{pr}} \text{Var}\, G_{\text{pr}} \left(-C_{\text{pr}} + B_{\text{res}\to\text{pr}} R_{\text{pr}} + D_{\text{pr}} S_{\text{pr}}\right)$$
$$+ \delta \pi_{\text{res}} \text{Var}\, G_{\text{res}} \left(-C_{\text{res}} + B_{\text{pr}\to\text{res}} R_{\text{res}} + D_{\text{res}} S_{\text{res}}\right) \tag{A.9}$$

as shown in Taylor (1990).

We can study the expression in Eq. (A.9) more easily if we assume that roles in the Ultimatum game are assigned randomly. In that case, we have no reason to think that the reproductive value of partners differ, nor do we have reason to believe that genetic variance, relatedness, or synergy depends on role (e.g. as in Hawk–Dove–Bourgeois game between relatives studied by Maynard Smith, 1982). Given the assumption, then, we obtain

$$\text{Cov}(W, G) \propto -\frac{C_{\text{pr}} + C_{\text{res}}}{2} + \frac{B_{\text{res}\to\text{pr}} + B_{\text{pr}\to\text{res}}}{2} R + \frac{D_{\text{pr}} + D_{\text{res}}}{2} S, \tag{A.10}$$

where the multiplicative factor of one-half appears because an individual is equally likely to be proposer or responder, and where

$$-C_{\text{pr}} = \frac{1}{2} - (1 - \varepsilon), \quad -C_{\text{res}} = -\varepsilon, \quad B_{\text{pr}\to\text{res}} = \frac{1}{2} - \varepsilon, \quad B_{\text{res}\to\text{pr}} = -(1 - \varepsilon)$$

$$D_{\text{pr}} = \frac{1}{2} - (1 - \varepsilon) - \left(\frac{1}{2} - (1 - \varepsilon)\right) - (-(1 - \varepsilon)) = 1 - \varepsilon$$

$$D_{\text{res}} = \frac{1}{2} - \varepsilon - (-\varepsilon) - \left(\frac{1}{2} - \varepsilon\right) = \varepsilon. \tag{A.11}$$

The reader will benefit from a more detailed explanation of the expressions in line (A.11). We obtain $-C_{\text{pr}}$ by determining the change in payoff experienced by a focal individual in the role of proposer who decides to change from rational to fair (i.e., from a to A), given that they are paired with a rational responder (see line (A.3)). Because this focal individual would have received $(1 - \varepsilon)$ (had it been rational) but now receives $\frac{1}{2}$ (it has become fair), we determine $-C_{\text{pr}}$ to be $\frac{1}{2} - (1 - \varepsilon)$. We obtain $-C_{\text{res}} = 0 - \varepsilon$ in an analogous way, but we do so by placing the focal individual in the role of responder.

We obtain $B_{\text{pr}\to\text{res}}$ by determining the change in payoff experienced by a rational focal individual, in the role of responder, when the proposer with whom the individual is paired decides to change from rational to fair (see line (A.3)). Whereas the focal individual would have received ε (i.e., it would have responded by accepting any non-zero proposal), it now receives $\frac{1}{2}$ because its partner has switched its approach to the game. This leaves us with $B_{\text{pr}\to\text{res}} = \frac{1}{2} - \varepsilon$. We find $B_{\text{res}\to\text{pr}} = 0 - (1 - \varepsilon)$ using the same approach, but with the focal individual cast as proposer.

Determination of coefficients D_{pr} and D_{res} is more involved. The former is associated with the change in payoff experienced by focal individual, in the role of proposer, when both the focal individual and its partner switch from rational to fair (see line (A.3)). The correct D_{pr}, however, describes only the non-additive effects associated with the switch: the coefficients $-C_{\text{pr}}$ and $B_{\text{res}\to\text{pr}}$ have already completely captured the additive effects. It follows that if the total change in

the proposer's payoff resulting from the partners' simultaneous shift to fairness is $\frac{1}{2} - (1 - \varepsilon)$, then the correct value of D_{pr} is determined by subtracting $-C_{\text{pr}}$ and $B_{\text{res}\to\text{pr}}$ from $\frac{1}{2} - (1 - \varepsilon)$. We find D_{res} in the same way, but place the focal individual in the role of responder.

From (A.10) and (A.11) we find that, when roles are decided at random,

$$\text{Cov}(W, G) \propto -\frac{1 + R}{2} + S. \tag{A.12}$$

In this case, when $S < 1/2$, meaning fair individuals are more likely to be partnered with rational individuals, fairness itself can be favoured only when $R < 0$, meaning partners are non-kin (Fig. A.1). When $S > 1/2$, fair individuals tend to be partnered with fair individuals, and so fairness can be favoured when R over a range that includes both negative and positive values, meaning partners can be non-kin or kin (Fig. A.1).

Appendix B. Actor-centred model

The expression in (A.9) takes a neighbour-modulated approach to fitness accounting. For us, this means that (A.9) focuses our attention on an individual in role j and on how it is affected by itself and its partner. In the main text, however, we take an actor-centric approach to fitness accounting and enumerate how an individual in role j affects itself and its partner. To do this, we re-write (A.9) as,

$$\begin{aligned}
\text{Cov}(W, G) &= \delta \pi_{\text{pr}} \text{Var}\, G_{\text{pr}} \left(-C_{\text{pr}} + D_{\text{pr}} S_{\text{pr}}\right) + \pi_{\text{pr}} B_{\text{res}\to\text{pr}} (f - p_{\text{pr}} p_{\text{res}}) \\
&\quad + \delta \pi_{\text{res}} \text{Var}\, G_{\text{res}} \left(-C_{\text{res}} + D_{\text{res}} S_{\text{res}}\right) \\
&\quad + \pi_{\text{res}} B_{\text{pr}\to\text{res}} (f - p_{\text{pr}} p_{\text{res}}) \\
&= \delta \pi_{\text{pr}} \text{Var}\, G_{\text{pr}} \left(-C_{\text{pr}} + D_{\text{pr}} S_{\text{pr}}\right) + \text{Var}\, G_{\text{res}} \pi_{\text{pr}} B_{\text{res}\to\text{pr}} R_{\text{res}} \\
&\quad + \delta \pi_{\text{res}} \text{Var}\, G_{\text{res}} \left(-C_{\text{res}} + D_{\text{res}} S_{\text{res}}\right) \\
&\quad + \text{Var}\, G_{\text{pr}} \pi_{\text{res}} B_{\text{pr}\to\text{res}} R_{\text{pr}} \\
&= \delta \text{Var}\, G_{\text{pr}} \underbrace{\left(-C_{\text{pr}} \pi_{\text{pr}} + R_{\text{pr}} B_{\text{res}\to\text{pr}} \pi_{\text{res}} + D_{\text{pr}} S_{\text{pr}} \pi_{\text{pr}}\right)}_{\text{(a)}} \\
&\quad + \delta \text{Var}\, G_{\text{res}} \underbrace{\left(-C_{\text{res}} \pi_{\text{res}} + R_{\text{res}} B_{\text{res}\to\text{pr}} \pi_{\text{pr}} + D_{\text{res}} S_{\text{res}} \pi_{\text{res}}\right)}_{\text{(b)}}.
\end{aligned} \tag{B.1}$$

By re-organizing terms in the actor-centred way we can more easily identify the relevant pieces in cases where the locus in question either

(a) controls proposer behaviour only (relevant piece labelled (a) in Eqs. (B.1)), or
(b) controls responder behaviour only (relevant piece labelled (b) in Eqs. (B.1)).

In situation (a), proposers carrying the A allele always make fair proposals, but responders carrying the A allele accept any non-zero proposal with which they are presented. In situation (b), responders carrying the A allele always reject unfair offers, but proposers carrying the A allele make unfair offers.

One key consequence of conditional behaviour is that the synergistic terms in (A.11) are lost. Loss of synergy has also been observed in other games following the introduction of conditional behaviour (Queller, 1996). For (a), when only the proposer acts we have

$$-C_{\text{pr}} = \frac{1}{2} - (1 - \varepsilon), \quad -C_{\text{res}} = 0, \quad B_{\text{pr}\to\text{res}} = \frac{1}{2} - \varepsilon, \quad B_{\text{res}\to\text{pr}} = 0$$

$$D_{\text{pr}} = \frac{1}{2} - (1 - \varepsilon) - \left(\frac{1}{2} - (1 - \varepsilon)\right) - (0) = 0$$

$$D_{\text{res}} = \frac{1}{2} - \varepsilon - (0) - \left(\frac{1}{2} - \varepsilon\right) = 0 \tag{B.2}$$

and we note that all terms in (b) have vanished. By contrast, for (b), when the only responder acts, we have

$$-C_{\text{pr}} = 0, \quad -C_{\text{res}} = -\varepsilon, \quad B_{\text{pr}\to\text{res}} = 0, \quad B_{\text{res}\to\text{pr}} = -(1 - \varepsilon)$$
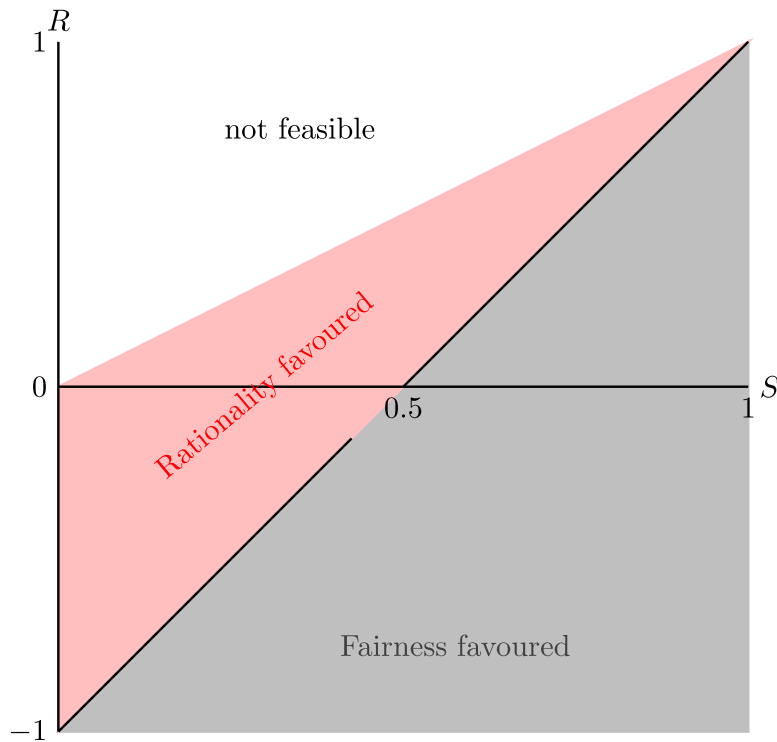
Fig. A.1. When roles are assigned at random, fairness is favoured over rationality only when the coefficient of relatedness between partners, R, and the coefficient of synergism, S, take specific values highlighted in grey. The line separating the grey region from the red region is $R = 2S - 1$. The region labelled as non-feasible corresponds to $R > S$ which cannot occur.

$$D_{\mathrm{pr}} = 0 - (1 - \varepsilon) - (0 - (1 - \varepsilon)) - (0) = 0$$

$$D_{\mathrm{res}} = -\varepsilon - (-\varepsilon) - (0) = 0. \tag{B.3}$$

and now all terms in (a) have vanished. To help the reader, we emphasize that the coefficients in lines (B.2) and (B.3) are determined using the same basic steps outlined in the previous appendix. The key difference, here, is that fairness cannot expressed by one of the two partners: in (B.2) the responder is always rational, while in (B.3) the proposer is always rational. These constraints give rise to the zeros that we see in (B.2) and (B.3), respectively. We also emphasize that, in the main text, we weight C_j and $B_{\neg j \to j}$ by $\pi_j = 1 - \pi_{\neg j}$ to obtain payoffs B, C, and D in lines (4) (scenario a) and line (5) (scenario b).

Appendix C. Dynamics

We derived Eq. (A.9) by assuming δ is small, i.e. by assuming the social interaction between proposer and responder contributes only very little to the overall fitness of each. This same assumption also allows us to estimate π_j, $\mathrm{Var}\,G_j$, R_j, and S_j by calculating each when $\delta = 0$. The ability to estimate key quantities is only useful if we have a particular dynamic scenario in mind, and we turn our attention to dynamics in this section.

We assume the population is finite, is not subject to selection ($\delta = 0$), but is subject to drift. We assume further that the two alleles at the locus in question, a and A, are maintained in the long term by infrequent symmetric mutation. In fact, for us, mutations are so rare that we can (have) neglected them elsewhere in our analysis.

From the description above we can conclude that quantities like variance, relatedness, and synergy will wander over time in a random manner. Ideally, we would average out the randomness so that predictions in the main text reflect our expectations in some sense. The averaging we choose, here, takes expectations over all realizations of the random process described above, in line with previous work (e.g. Taylor et al., 2007). By symmetry, it is clear that the frequency of the

A allele among individuals in any role is one-half; thus averaging leads us to $p_{\mathrm{pr}} = p_{\mathrm{res}} = 1/2$.

No further discussion is needed, but we will add a few more comments. First, we note that we use $\mathrm{Var}\,G_j = p_j(1 - p_j)$. Whether this is reasonable or not is of no consequence since the variance is just a constant term we factor out of the covariance expression. Second, we leave π_j unspecified but note that it could also be estimated provided additional assumptions about population dynamics are made. Third, this section of the appendix is only relevant to the case in which both proposer and responder express fairness, because it is only in this case that we have to contend with both relatedness coefficients and coefficients of synergy. On that topic, we find that $S = f/(1/2)$ and so

$$R = \frac{(1/2)S - (1/2)^2}{(1/2)^2} \quad \leftrightarrow \quad \frac{1 + R}{2} = S. \tag{C.1}$$

This leads to Eq. (2). Importantly, it also suggests that the result presented in line (A.12) is zero. In other words, when roles are determined at random fairness is not visible to selection, at least approximately. We might use a higher-order approximation to improve on (A.12), but that is beyond the scope of this work. That said, we focus instead on cases where roles are not determined at random, so $\pi \neq 1/2$ and the population is truly heterogeneous.

References

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., List, J.A., 2011. Stakes matter in ultimatum games. Am. Econ. Rev. 101, 3427–3439. http://dx.doi.org/10.1257/aer.101.7.3427.

Boccia, M.L., Laudenslager, M., Reite, M., 1988. Food distribution, dominance, and aggressive behaviors in bonnet macaques. Am. J. Primatol. 16, 123–130. http://dx.doi.org/10.1002/ajp.1350160203.

Cameron, L.A., 1999. Raising the stakes in the ultimatum game: experimental evidence from Indonesia. Econ. Inq. 37, 47–59. http://dx.doi.org/10.1111/j.1465-7295.1999.tb01415.x.

Cooper, G.A., Levin, S.R., Wild, G., West, S.A., 2018. Modeling relatedness and demography in social evolution. Evol. Lett. 2, 260–271. http://dx.doi.org/10.1002/evl3.69.

de Waal, F.B.M., 1986. The integration of dominance and social bonding in primates. Q. Rev. Biol. 61, 459–479. http://dx.doi.org/10.1086/415144.

Eisenbruch, A.E., Grillot, R.L., Maestripieri, D., Roney, J.R., 2016. Evidence of partner choice heuristics in a one-shot bargaining game. Evol. Hum. Behav. 37, 429–439. http://dx.doi.org/10.1016/j.evolhumbehav.2016.014.002.

Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425, 785–791. http://dx.doi.org/10.1038/nature02043.

Gardner, A., 2017. The purpose of adaptation. Interface Focus 7, 20170005. http://dx.doi.org/10.1098/rsfs.2017.0005.

Gardner, A., West, S.A., Wild, G., 2011. The genetical theory of kin selection. J. Evol. Biol. 24, 1020–1043. http://dx.doi.org/10.1111/j.1420-9101.2011.02236.x.

Geritz, S.A.H., Kisdi, E., 2000. Adaptive dynamics in diploid, sexual populations and the evolution of reproductive isolation. Proc. R. Soc. Lond. B 267, 1671–1678. http://dx.doi.org/10.1098/rspb.2000.1194.

Gintis, H., 2000. Strong reciprocity and human sociality. J. Theoret. Biol. 206, 169-179. http://dx.doi.org/10.1006/jtbi.2000.2111.

Grafen, A., 1984. Natural selection, kin selection and group selection. In: Krebs, J., Davies, N. (Eds.), Behavioural Ecology, second ed. Oxford, Blackwell, pp. 62–84, chapter 3.

Grafen, A., 1985a. Evolutionary theory: Hamilton's rule OK. Nature 318, 310–311. http://dx.doi.org/10.1038/318310a0.

Grafen, A., 1985b. A geometric view of relatedness. Oxf. Surv. Evol. Biol. 2, 28–90.

Grafen, A., 2015. Biological fitness and the Price equation in class-structured populations. J. Theoret. Biol. 373, 62–72. http://dx.doi.org/10.1016/j.jtbi.2015.02.014.

Grafen, A., 2018. The left hand side of the fundamental theorem of natural selection. J. Theoret. Biol. 456, 175–189. http://dx.doi.org/10.1016/j.jtbi.2018.07.022.

Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J. Econ. Behav. Org. 3, 367–388. http://dx.doi.org/10.1016/0167-2681(82)90011-7.

Hamilton, W.D., 1964. The genetical evolution of social behaviour. I and II. J. Theoret. Biol. 7, 1–52. http://dx.doi.org/10.1016/0022-5193(64)90038-4.

Härdling, R., 2007. Fairness evolution in the ultimatum game is a function of reward size. J. Theoret. Biol. 246, 720–724. http://dx.doi.org/10.1016/j.jtbi.2007.01.026.

Killingback, T., Studer, E., 2001. Spatial Ultimatum Games, collaborations and the evolution of fairness. Proc. R. Soc. B 268, 1797–1801. http://dx.doi.org/10.1098/rspb.2001.1697.

Lehmann, L., Feldman, M.W., Foster, K.R., 2008. Cultural evolution can inhibit the evolution of altruism. Amer. Nat. 172, e12–e24. http://dx.doi.org/10.1086/587851.

Lehmann, L., Rousset, F., Roze, D., Keller, L., 2007. Strong reciprocity or strong ferocity? a population genetic view of the evolution of altruistic punishment. Amer. Nat 170, 21–36. http://dx.doi.org/10.1086/518568.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge, UK.

Michod, R.E., Hamiton, W.D., 1980. Coefficients of relatedness in sociobiology. Nature 288, 694–697. http://dx.doi.org/10.1038/288694a0.

Milinski, M., 2013. Chimps play fair in the ultimatum game. Proc. Natl. Acad. Sci. USA 110, 1978–1979. http://dx.doi.org/10.1073/pnas.1222132110.

Page, K.M., Nowak, M.A., 2000. A generalized adaptive dynamics framework can describe the evolutionary Ultimatum Game. J. Theoret. Biol. 209, 173–179. http://dx.doi.org/10.1006/jtbi.2000.2251.

Price, G.R., 1970. Selection and covariance. Nature 227, 520–521. http://dx.doi.org/10.1038/227520a0.

Proctor, D., Williamson, R.A., de Waal, S.F., 2013. Chimpanzees play the ultimatum game. Proc. Natl. Acad. Sci. USA 110, 2070–2075. http://dx.doi.org/10.1073/pnas.1220806110.

Queller, D.C., 1985. Kinship, reciprocity and synergism in the evolution of social-behaviour. Nature 318, 366–367. http://dx.doi.org/10.1038/318366a0.

Queller, D.C., 1996. The measurement and meaning of inclusive fitness. Anim. Behav. 51, 229–232. http://dx.doi.org/10.1006/anbe.1996.0020.

Queller, D.C., 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. Proc. Natl. Acad. Sci. USA 108, 10792–10799. http://dx.doi.org/10.1073/pnas.1100298108.

Sánchez-Amaro, S., Call, J., Tomasello, M., 2018. Chimpanzees' understanding of social leverage. PLOS ONE 13, e0207868. http://dx.doi.org/10.1371/journal.pone.0207868.

Sapolsky, R.M., 2005. The influence of social hierarchy on primate health. Science 308, 648–652. http://dx.doi.org/10.1126/science.1106477.

Taylor, P.D., 1988. Inclusive fitness models with two sexes. Theor. Popul. Biol. 34, 145–168. http://dx.doi.org/10.1016/0040-5809(88)90039-1.

Taylor, P.D., 1990. Allele-frequency change in a class-structured population. Amer. Nat. 135, 95–106. http://dx.doi.org/10.1086/285034.

Taylor, P.D., 2017. Inclusive fitness in finite populations—effects of heterogeneity and synergy. Evolution 71, 508–525. http://dx.doi.org/10.1111/evo.13160.

Taylor, P.D., Day, T., Wild, G., 2007. From inclusive fitness to fixation probability in homogeneous structured populations. J. Theoret. Biol. 249, 101–110. http://dx.doi.org/10.1016/j.jtbi.2007.07.006.

Watts, D.P., 2010. Dominance, power, and politics in nonhuman and human primates. In: Kappeler, P.M., Silk, J. (Eds.), Mind the Gap: Tracing the Origins of Human Universals. Springer, Berlin, Heidelberg, pp. 109–138. http://dx.doi.org/10.1007/978-3-642-02725-3_5.

West, S.A., Griffin, A.S., Gardner, A., 2008. Social semantics: how useful has group selection been? J. Evol. Biol. 21, 374–385. http://dx.doi.org/10.1111/j.1420-9101.2007.01458.x.

Wild, G., Traulsen, A., 2007. The different limits of weak selection and the evolutionary dynamics of finite populations. J. Theoret. Biol. 247, 382–390. http://dx.doi.org/10.1016/j.jtbi.2007.03.015.

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., Simunovic, D., 2012. Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. Proc. Natl. Acad. Sci. USA 109, 20364–20368. http://dx.doi.org/10.1073/pnas.1212126109.

Zeng, T.C., Cheng, J.T., Henrich, J., 2022. Dominance in humans. Philos. Trans. R. Soc. B 377, 20200451. http://dx.doi.org/10.1098/rstb.2020.0451.