1. Introduction

Various comments and TODOs:

- Discrepancy between the literature on groups, including intergroup conflict parochial altruism, and the experimental literature. The literature considers rich groups, with layers of intra and intergroup interactions. But the experiments show that minimal groups trigger group psychology. We offer an evolutionary explanation for how group cognition can evolve out of a minimal group setup. (Need to check this is fair to the theoretical lit)
- Possibly field evidence from segmented societies: me and my brother against my cousin, me and my cousin against the rest of the clan, etc. Shows that the psychology is flexible. (DONE)
- Many models assume that the cognitive machinery is already there.
- Mention that the "group structure"/"common fate" explanation is common sense and may explain some group cooperation.
- Mustn't claim our model explains everything in real world conflicts
- Model can even explain the evol. of group categorization (i.e. cognition), since selfish types don't need to even observe group labels.
- In discussion: why can't a direct reciprocator outcompete? Possible answers: can't individuate outgroup members (cf Fearon Laitin); don't regularly interact with any one individual.

Humans categorize people into groups, hold attitudes and stereotypes towards outgroups, and discriminate against them (Brewer and Kramer, 1985; Neuberg and Schaller, 2008). They also reciprocate towards groups, targeting bystanders for the actions of their fellow group members (Gaertner, Iuzzini, and O'Mara, 2008; Hugh-Jones and Leroch, 2017; Hugh-Jones, Ron, and Zultan, 2019; Lickel et al., 2006; Stenstrom et al., 2008). This group psychology may feed into intergroup conflicts between clans, ethnicities and nations (Haushofer, Biletzki, and Kanwisher, 2010; Horowitz, 1985, 2001). While the psychological mechanisms associated with group behavior and cognition have been exten-

sively studied, the evolutionary origins of group psychology remain obscure (De Dreu, Gross, and Romano, 2023).

This paper describes an evolutionary model of intergroup reciprocity. We show that group psychology can evolve even in an environment with a very minimal group structure. Our model has an arbitrary number of groups. Groups do not determine the interaction structure: everyone interacts with everyone else. Nor do they affect mating or evolution: mating is random, fitness affects reproduction across the whole population, and children are rematched randomly into new groups. The cognitive requirements of the model are also quite minimal: Individuals can learn other individuals' group membership, and they can recall others' past behaviour towards them. In short, groups are just arbitrary labels.

As it turns out, this minimal structure is enough for intergroup reciprocity to evolve in a finitely repeated prisoner's dilemma. Group-reciprocal types, who cooperate with someone from group G based on how many people from group G previously cooperated with them, can have higher fitness. The equilibrium features a mix of group-reciprocal and selfish types.

Group reciprocity involves three capabilities: recording an outgroup's "image" (its members' average level of previous cooperation), categorizing individuals as members of this outgroup, and acting towards them on the basis of these two pieces of information. So, group reciprocity can be a basis for the evolution of outgroup categorization, outgroup "image scoring", and discrimination.

The capacity for group reciprocity may have been a stepping stone in the evolution of human cooperation at scale. Interethnic peace may be sustained by the threat of reciprocation. This could have allowed humans to live in peace with neighbouring groups (perhaps punctuated by episodes of punitive conflict), whereas chimpanzees always attack neighbouring bands on contact (Wrangham and Glowacki, 2012). Group reciprocity could also enable deeper forms of intergroup cooperation, such as trade.

Here is the logic behind our result. Groups of group-reciprocators establish cooperative relationships with other groups of group-reciprocators, while they defect against groups of selfish types. So, when some people are group

reciprocators, individual fitness depends on one's group's reputation. Group reciprocators contribute to this reputation, which is a group-level public good. Thus being a group reciprocator is good for the group but costly for the individual. This then gives rise to a standard intrademic group selection model, where the group-level advantages balance out the individual-level costs (Wade, 1978; Wilson, 1983). Selfish free-riders within a group of mostly group-reciprocators get the highest payoff, but this is balanced out because most selfish types are in mostly-selfish groups, who have too few reciprocators to sustain intergroup cooperation.

We use computations to check the robustness of our theoretical model. Our results confirm that even with relatively rare interactions, group reciprocity can evolve.

Our model also suggests some insights about how group reciprocal motivations are shaped by evolutionary pressure. The most robust forms of group reciprocity have high thresholds of cooperation - that is, reciprocators only cooperate with groups a high proportion of whose members previously cooperated towards them. These high thresholds make it less likely that selfish free riders will invade, since only groups which are almost all reciprocators get the benefit of mutual cooperation. The resulting equilibrium is fragile: it sustains a high level of cooperation, but is also sensitive to small amounts of defection.

2. Literature review

Evidence for group reciprocity, or "vicarious revenge", comes from the lab (Gaertner, Iuzzini, and O'Mara, 2008; Hugh-Jones and Leroch, 2017; Hugh-Jones, Ron, and Zultan, 2019; Lickel et al., 2006; Romano, Saral, and Wu, 2022; Stenstrom et al., 2008) and the field, including violent intergroup conflict (Haushofer, Biletzki, and Kanwisher, 2010) and feuding institutions (Boehm, 1984; Chagnon, 1988).

Theories of the evolution of group reciprocity fall into two clusters: work done by evolutionary psychologists, and formal models. Evolutionary psychologists argue that groups are coalitions driven by exchange of benefits (Kurzban

and Neuberg, 2015; Tooby and Cosmides, 2010). The theory has generated novel predictions, including that men will show stronger group psychology than women, and that disgust reactions to outgroups may be linked to disease avoidance (Van Vugt and Park, 2009). Less work has been done to understand why different outgroups elicit different reactions, but the basic prediction is that group psychology will be activated for groups that are perceived as potential coalitions (Kurzban, Tooby, and Cosmides, 2001). One issue for these theories is explaining the *flexibility* of group psychology (Dunham, 2018). In the lab, even "minimal" groups can evoke group psychology (Tajfel et al., 1971), while categorization by race can be overruled by other markers of coalition membership (Kurzban, Tooby, and Cosmides, 2001). In the field, segmented societies feature flexible coalitions, in which minor lineages divide over local issues but unite to oppose rival clans (Fortes and Evans-Pritchard, 2015). Columbus et al. (2023) argue informally that group reciprocity can be supported by reputation mechanisms at group and individual level. Groups have an interest in appearing "tough", i.e. able to reciprocate harms; individuals who reciprocate on behalf of their group may gain an individual reputation for toughness and (parochial) prosociality.

Formal evolutionary models have focused on "parochial altruism" (Choi and Bowles, 2007; Rusch, 2014; Smirnov et al., 2007). These models typically feature only two groups, and do not aim to explain why different outgroups are treated differently. They also use "strong" groups: assumptions about ingroup cooperation, intergroup conflict, group-relative fitness and population structure are built into the model.

Fearon and Laitin (1996) use an infinitely repeated game to explain how different ethnic groups can live at peace. In their "spiral regime", defection by any member of ethnic group A towards a member of B leads to subsequent defection by all members of B towards members of A. The goal is to explain institutions that support interethnic cooperation, such as feuds. Our theory has a different setup and motivation. We examine the evolutionary stability of different types in a finite game. In Fearon and Laitin (1996), cooperation is supported by the threat of collective punishment. Here, group reciprocity is evolutionarily stable because individual free-riding is balanced by group selec-

4

tion. So, our model is designed to explain the evolution of "strong" reciprocal motivations in humans (Gintis, 2000), rather than the stability of institutions supporting interethnic peace.

Here, we show that intergroup psychology can evolve even under "minimal" conditions: groups are arbitrary labels, with no special mechanisms for in-group cooperation, collaboration, or resource sharing, no within-group mating and no explicit group-level conflict. Even under these conditions, groups who act reciprocally towards other groups can establish intergroup cooperation and gain a fitness advantage.

We suggest that group reciprocal behaviour involves multiple faculties: categorizing people as members of a given outgroup; combining members' actions into a group-level "image score"; motivations to respond reciprocally to outgroup members on the basis of the score. By providing a pathway for these phenomena to evolve, group reciprocity can explain why humans hold intergroup attitudes (Brewer and Kramer, 1985; Kurzban, Tooby, and Cosmides, 2001). One simple descriptive framework is that groups, like individuals, are perceived on the two dimensions of warmth and competence (Fiske, Cuddy, and Glick, 2007). Our model captures the warmth dimension. It also suggests why the propensity to hold beliefs about outgroups may be evolutionarily stable: when humans are group-reciprocators, aggregated group characteristics (e.g. how often that group has cooperated with your group, and vice versa) are important in predicting individual behaviour. There is an analogy here with the literature on individual indirect reciprocity. Models of "image scoring" or "standing", in which individuals keep track of other individuals' standing (e.g. Nowak and Sigmund, 2005), can explain the evolution of moral norms and reputations. The same holds here, but at group level.

Group reciprocity is a form of "upstream reciprocity", where an individual who is helped or harmed by someone in turn helps or harms other third parties (Boyd and Richerson, 1989). It is hard for upstream reciprocity to evolve, because it does not naturally lead to stable bilateral relationships (Nowak and Roch, 2007).[1] Indeed, there is only mixed experimental evidence for upstream

---

[1]By contrast, we have satisfying theoretical explanations for why group membership might matter for "downstream" or "indirect" reciprocity, where people help someone who previously

5

reciprocity (Ben-Ner et al., 2004; Greiner and Levati, 2005; Horita et al., 2016; Stanca, 2009; van Apeldoorn and Schram, 2016). We show that the existence of groups makes this problem easier, by allowing different *groups* to form cooperative, fitness-increasing bilateral relationships. Some observed cases of upstream reciprocity (e.g. Mujcic and Leibbrandt, 2018; Yuan et al., 2019) may involve elements of group reciprocity. For sure, group reciprocity provides many of the best-known examples of upstream reciprocity, like violent intergroup revenge.

3. Model

We consider a mixed population of size N that is composed of two types, selfish and group reciprocators (GR). At the beginning of each generation, the population randomly divides into ~~a large number of~~ G groups of size n each (so that $N = nG$). Let p denote the population share of GR types. Let ~~p_g be the proportion~~ l_g be the number of GR in group g ~~, which is distributed binomially~~.

At every step t, everybody interacts with everybody. In each pair, each individual chooses between cooperation and defection. Cooperation entails a cost c to the cooperator and a benefit b to her partner. Defection carries no costs. That is, each pair plays the following Prisoner's Dilemma game:

	Cooperate	Defect
Cooperate	$b - c$	$-c$
Defect	b	0

Selfish types always defect. A GR individual i starts by cooperating, and then cooperates with all individuals belonging to group g with a probability $\phi(l_{gi})$, where l_{gi} is the ~~proportion~~ number of individuals from group g who cooperated with individual i in round $t - 1$. $\phi(\cdot)$ is monotonically weakly increasing. We consider the cutoff strategy:

$$\phi(l_{gi}) = \begin{cases} 1 & \text{if } l_{gi} \geq k \\ 0 & \text{otherwise.} \end{cases}$$

helped others.

6

Equilibrium in this game is as follows: in period 1, group reciprocators help everyone. In period 2, all group reciprocators help those in "supraliminal" groups with $l_g \geq k$. In periods 3 and above, group reciprocators in supraliminal groups help those in supraliminal groups; they defect against everyone else, and all other individuals defect.

The fitness is the payoff at the limit where $t \to \infty$. Equivalently, since the game always settles to a stationary action profile, it is the average payoff of T rounds when $T \to \infty$.

Individuals' fitness therefore depends only on whether they are in a "supraliminal" group, and on their type. Let q be the proportion of supraliminal groups. Let $\bar{p}$ be the proportion of GR individuals in supraliminal groups (out of the total population in such groups). Let $\underline{p}$ be the proportion of GR individuals in subliminal groups (out of the total population in such groups). It follows that

- Group reciprocators in supraliminal groups get a payoff of $(\bar{p}qb - qc)N$ $= \Sigma_{\{g|l_g \geq k\}}[l_g b - nc]$.
- Selfish types in supraliminal groups get $\bar{p}qbN = \Sigma_{\{g|l_g \geq k\}}l_g b$.
- Group reciprocators and selfish types in subliminal groups get 0.

After each generation, reproductive success is proportional to fitness, the total population size stays the same, and children are remixed randomly into new groups of the same size. The mean fitness of the GR type is

$$\frac{\bar{p}q(q(\bar{p}b - c))N}{\underline{p}} = \frac{\Sigma_{\{g|l_g \geq k\}}l_g \Sigma_{\{g|l_g \geq k\}}[l_g b - nc]}{\Sigma_g l_g}$$

and the mean fitness of selfish types is

$$\frac{(1 - \bar{p})q(q\bar{p}b)N}{1 - \underline{p}} = \frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g)\Sigma_{\{g|l_g \geq k\}}l_g b}{\Sigma_g(n - l_g)}$$

After rearranging, the mean fitness of reciprocators is higher if

$$\frac{\bar{p} - \underline{p}}{1 - \underline{p}} \geq \frac{c}{b} \Longleftrightarrow \frac{N\Sigma_{\{g|l_g \geq k\}}l_g - M\Sigma_g l_g}{N - \Sigma_g l_g} \geq M\frac{b}{c}. \tag{1}$$

where $M \equiv qN = \Sigma_{\{g|l_g \geq k\}}n$ and

7

$$\bar{p} = \frac{\Sigma_{\{g|l_g \geq k\}} l_g}{M}$$

Proposition 1. *If $M \neq \emptyset$, then the LHS of equation* (**??**) *is decreasing in* $p = \frac{\Sigma_g l_g}{N}$ ~~and is equal to the threshold k when $p = 0$.~~[2]

Proof. See Appendix.

First rewrite the LHS of (**??**) as follows: $LHS = M - Nx$, where $x \equiv \frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g)}{N - \Sigma_g l_g}$ is the proportion of selfish types who belong to supraliminal groups. A marginal increase in p is modeled by a random selection of a selfish type who is replaced by a GR-type. That is, with probability x this selfish type belongs to a supraliminal group and otherwise it belongs to a subliminal group.

Suppose first that the group from which the selfish type is randomly selected cannot switch from being subliminal to being supraliminal (even if was of size $l_g = k - 1$).[3] Then, with probability x, the increase in the LHS is given by $-N(\frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g) - 1}{N - \Sigma_g l_g - 1} - x)$, and with probability $1 - x$ the increase in the LHS is given by $-N(\frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g)}{N - \Sigma_g l_g - 1} - x)$. Therefore, the expected increase in the LHS is given by

$$-N[x(\frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g) - 1}{N - \Sigma_g l_g - 1} - x) + (1 - x)(\frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g)}{N - \Sigma_g l_g - 1} - x)]$$

$$= -N[\frac{x[\Sigma_{\{g|l_g \geq k\}}(n - l_g) - 1] + (1 - x)\Sigma_{\{g|l_g \geq k\}}(n - l_g)}{N - \Sigma_g l_g - 1} - x]$$

$$= -N[\frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g) - x}{N - \Sigma_g l_g - 1} - x].$$

---

[2] ~~When the share of GR in the population is very small, the probability that $p_g = k$ conditional on $p_g \geq k$ goes to one.~~

[3] What we mean by that is that this group will not be included in the sum $\Sigma_{\{g|l_g \geq k\}}$ even if it becomes of size k.

Recalling that $x = \frac{\Sigma_{\{g|l_g \geq k\}}(n-l_g)}{N - \Sigma_g l_g}$ we get that

$$-N\left[\frac{\Sigma_{\{g|l_g \geq k\}}(n - l_g) - x}{N - \Sigma_g l_g - 1} - x\right] = -N\left[\frac{x(N - \Sigma_g l_g) - x}{N - \Sigma_g l_g - 1} - x\right] = 0.$$

That is, if we do not allow groups to switch from being subliminal to being supraliminal, then the effect of randomly replacing a selfish type with a GR-type on the LHS is zero. However, if this replacement results in a subliminal group becoming supraliminal (note that the opposite change is impossible), then x, the proportion of selfish types who belong to supraliminal groups, necessarily increases (there will be more selfish types in supraliminal groups out of less selfish types in the population), implying that $M - Nx$ (the LHS) must decrease. $\square$

It follows that there is a unique ESS with a positive share of group reciprocators if and only if $k > \frac{c}{b}$. Otherwise the population is homogeneously selfish in the unique ESS.

4. Simulations

The model includes two simplifying assumptions: a large number of groups, and a large number of rounds T in each generation. To check how these simplifications affect the result, we run simulations. We use the following sets of parameters: $c/b = 0.2, 0.5$; $T = 10, 20$; N groups = 10, 50; $k = 0.4, 0.8$. This gives 16 unique combinations of parameters. We always set $G = 8$ and start with half the population as GR types. Fitness is normalized so that the "sucker's payoff" $-c$ equals zero. For each combination we run 20 experiments, each lasting 500 generations. After 500 generations we record the proportion of group reciprocators.

Figure **??** shows the mean proportion of GR types across all experiments for each combination of parameters. For all combinations of parameters but one, either GR types or selfish types always went to fixation within 500 generations. In these cases, the proportion of GR types is just the proportion of experiments

9

where GR types went to fixation.

A shorter number of periods T leads to fewer GR types, as expected since it gives more weight to the early-round losses of GR types. The number of groups does not seem to affect the evolution of GR types. As expected, a higher cost-benefit ratio c/b leads to fewer GR types. A higher threshold k also usually leads to more GR types, and in particular when $k < c/b$, group reciprocity never evolves. One exception is with 50 groups and $T = 20$: in this case, more group reciprocators evolve when $k = 0.4$. This is also the sole parameter combination where not all experiments went to fixation.

In our analytic result with many groups, group size G doesn't matter. With finite numbers of groups, all groups may fall below the k-threshold in a given generation. This becomes more likely when G is large and when the population proportion of GR types is below the threshold (because when G is large, the proportion of GR types in each group becomes concentrated on the population proportion). So, group size may matter. We ran one further parameter set to check whether group reciprocity could be stable in larger groups. We set $G = 40$, N groups = 40, $T = 20$, $k = 0.6$ and $c/b = 0.1$. In 10 experiments, the proportion of group reciprocators was always between 60 and 70 per cent after 500 generations. So, evolution of group reciprocity is possible even with large groups, although it seems to require high benefit-cost ratios.

5. Conclusion

We introduced an evolutionary model of group reciprocity. Group reciprocal motivations can evolve under plausible parameter values for human populations [XXX can we say that?]. Evolved group reciprocity could explain why humans track group-level reputations. It also provides a simple way that upstream reciprocity can evolve.

The simplest alternative theory of group reciprocity is perhaps that many groups are able to solve collective action problems, i.e. to incent or persuade their members to maximize the payoff of the group. If so then reciprocity towards other groups could be optimal in the same way as individual-level reci-

Figure 1: Simulation results. Plots show proportions of group reciprocator types after 100 generations.
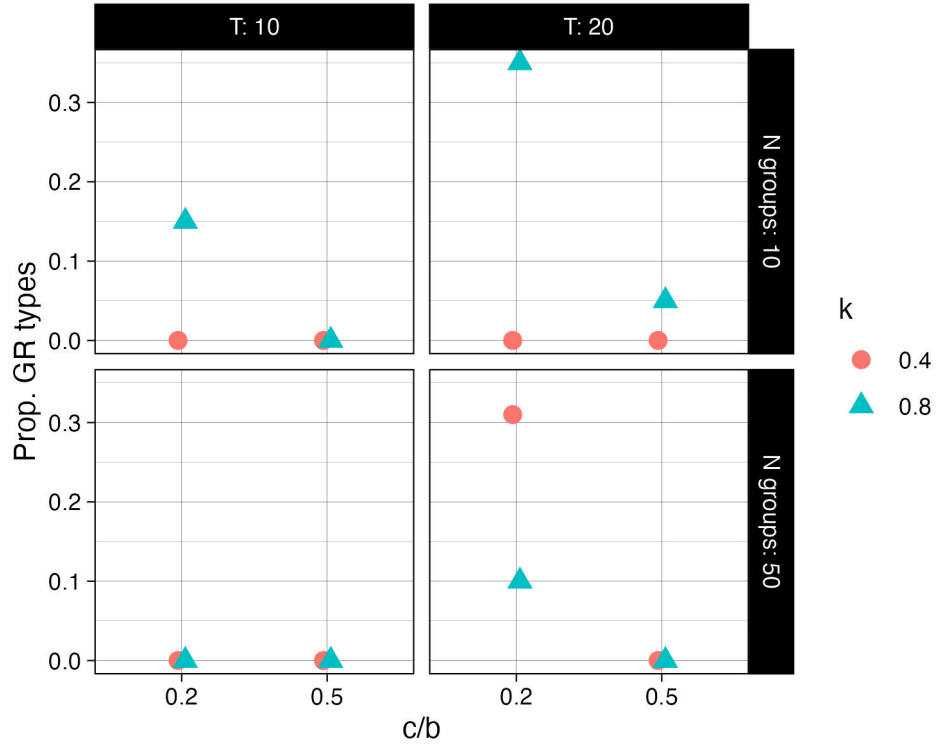
procity. Lab experiments show group reciprocity even among "minimal" groups which have no way to address collective action problems. But this might be because participants over-generalize, bringing into the lab their intuitions about behaviour that makes sense in the field. A useful next step would be to examine group reciprocity in the field, among groups which lack institutions for collective action.

References

Ben-Ner, Avner, Louis Putterman, Fanmin Kong, and Dan Magan (2004). Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization* 53.(3), pp. 333–352.

Boehm, Christopher (1984). Blood revenge: the enactment and management of conflict in Montenegro and other tribal societies. Philadelphia, PA: University of Pennsylvania Press.

Boyd, Robert and Peter J. Richerson (1989). The evolution of indirect reciprocity. *Soc Networks* 11.(3), pp. 213–236.

Brewer, Marilynn B. and Roderick M. Kramer (1985). The psychology of intergroup attitudes and behavior. *Annual review of psychology* 36.(1), pp. 219–243.

Chagnon, Napoleon A. (Feb. 1988). Life Histories, Blood Revenge, and Warfare in a Tribal Population. *Science* 239.(4843), pp. 985–992. URL: http://www.sciencemag.org/content/239/4843/985 (visited on 10/02/2013).

Choi, Jung-Kyoo and Samuel Bowles (2007). The coevolution of parochial altruism and war. *Science* 318.(5850), pp. 636–640.

Columbus, Simon, Isabel Thielmann, Ingo Zettler, and Robert Böhm (2023). Parochial reciprocity. *Evolution and Human Behavior*.

De Dreu, Carsten K. W., Jörg Gross, and Angelo Romano (2023). Group formation and the evolution of human social organization. *Perspectives on Psychological Science*.

Dunham, Yarrow (2018). Mere membership. *Trends in cognitive sciences* 22.(9), pp. 780–793.

Fearon, James D. and David D. Laitin (1996). Explaining interethnic cooperation. *American political science review* 90.(4), pp. 715–735.

Fiske, Susan T., Amy J. C. Cuddy, and Peter Glick (2007). Universal dimensions of social cognition: warmth and competence. *Trends in cognitive sciences* 11.(2), pp. 77–83.

Fortes, Meyer and Edward Evan Evans-Pritchard (2015). African political systems. Routledge.

Gaertner, Lowell, Jonathan Iuzzini, and Erin M. O'Mara (July 2008). When rejection by one fosters aggression against many: Multiple-victim aggression as a consequence of social rejection and perceived groupness. *J Exp Soc Psychol* 44.(4), pp. 958–970. URL: http://www.sciencedirect.com/science/article/pii/S0022103108000267 (visited on 07/20/2011).

Gintis, Herbert (2000). Strong reciprocity and human sociality. *Journal of theoretical biology* 206.(2), pp. 169–179.

Greiner, Ben and M. Vittoria Levati (2005). Indirect reciprocity in cyclical networks: an experimental study. *J Econ Psychol* 26.(5), pp. 711–731.

Haushofer, Johannes, Anat Biletzki, and Nancy Kanwisher (2010). Both sides retaliate in the Israeli–Palestinian conflict. *P Natl Acad Sci Usa* 107.(42), pp. 17927–17932.

Horita, Yutaka, Masanori Takezawa, Takuji Kinjo, Yo Nakawake, and Naoki Masuda (2016). Transient nature of cooperation by pay-it-forward reciprocity. *Scientific reports* 6.(1), p. 19471.

Horowitz, D. L. (1985). Ethnic Groups in Conflict. Berkeley: University of California Press.

— (2001). The deadly ethnic riot. University of California Press.

Hugh-Jones, David and Martin Alois Leroch (2017). Intergroup revenge: a laboratory experiment. *Homo Oeconomicus* 34, pp. 117–135.

Hugh-Jones, David, Itay Ron, and Ro'i Zultan (2019). Humans reciprocate by discriminating against group peers. *Evolution and Human Behavior* 40.(1), pp. 90–95.

Kurzban, Robert and Steven Neuberg (2015). Managing ingroup and outgroup relationships. *The handbook of evolutionary psychology*, pp. 653–675.

Kurzban, Robert, John Tooby, and Leda Cosmides (2001). Can race be erased? coalitional computation and social categorization. *Proceedings of the National Academy of Sciences* 98.(26), pp. 15387–15392.

Lickel, Brian, Norman Miller, Douglas M. Stenstrom, Thomas F. Denson, and Toni Schmader (Nov. 2006). Vicarious Retribution: The Role of Collective Blame in Intergroup Aggression. en. *Pers Soc Psychol Rev* 10.(4), pp. 372–390. URL: http://psr.sagepub.com/content/10/4/372 (visited on 06/10/2015).

Mujcic, Redzo and Andreas Leibbrandt (2018). Indirect reciprocity and prosocial behaviour: evidence from a natural field experiment. *The Economic Journal* 128.(611), pp. 1683–1699.

Neuberg, Steven L. and M. Schaller (2008). Intergroup prejudices and intergroup conflicts. *Foundations of evolutionary psychology* 401.

Nowak, Martin A. and Sébastian Roch (2007). Upstream reciprocity and the evolution of gratitude. *P Roy Soc B-biol Sci* 274.(1610), pp. 605–610.

Nowak, Martin A. and Karl Sigmund (2005). Evolution of indirect reciprocity. *Nature* 437.(7063), pp. 1291–1298.

Romano, Angelo, Ali Seyhun Saral, and Junhui Wu (2022). Direct and indirect reciprocity among individuals and groups. *Current Opinion in Psychology* 43, pp. 254–259.

Rusch, Hannes (2014). The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society of London B: Biological Sciences* 281.(1794), p. 20141539.

Smirnov, Oleg, Holly Arrow, Douglas Kennett, and John Orbell (2007). Ancestral war and the evolutionary origins of "heroism". *The Journal of Politics* 69.(4), pp. 927–940.

Stanca, Luca (2009). Measuring indirect reciprocity: whose back do we scratch? *Journal of Economic Psychology* 30.(2), pp. 190–202.

Stenstrom, Douglas M., Brian Lickel, Thomas F. Denson, and Norman Miller (Aug. 2008). The Roles of Ingroup Identification and Outgroup Entitativity in Intergroup Retribution. *Pers Soc Psychol B*. URL: `http://psp.sagepub.com/content/early/2008/08/20/0146167208322999` (visited on 06/10/2015).

Tajfel, Henri, Michael G. Billig, Robert P. Bundy, and Claude Flament (1971). Social categorization and intergroup behaviour. *European journal of social psychology* 1.(2), pp. 149–178.

Tooby, John and Leda Cosmides (2010). Groups in mind: the coalitional roots of war and morality. *Human morality and sociality: Evolutionary and comparative perspectives*, pp. 91–234.

Van Vugt, Mark and Justin H Park (2009). Guns, germs, and sex: how evolution shaped our intergroup psychology. *Social and Personality Psychology Compass* 3.(6), pp. 927–938.

Van Apeldoorn, Jacobien and Arthur Schram (2016). Indirect reciprocity; a field experiment. *PloS one* 11.(4), e0152076.

Wade, Michael J. (1978). A critical review of the models of group selection. *The Quarterly Review of Biology* 53.(2), pp. 101–114.

Wilson, David Sloan (1983). The group selection controversy: history and current status. *Annual review of ecology and systematics* 14.(1), pp. 159–187.

Wrangham, Richard W. and Luke Glowacki (2012). Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers. *Hum Nature* 23.(1), pp. 5–29.

Yuan, Yuan, Tracy Liu, Chenhao Tan, Qian Chen, Alex Pentland, and Jie Tang (2019). Gift contagion in online groups: evidence from WeChat red packets. *arXiv preprint arXiv:1906.09698.*

A. Proofs

A.1. Proof of Proposition ??

We will first prove an auxiliary lemma and then the main result.

Lemma 1. *Let $x_1, x_2, ..., x_n$ be iid Binomially-distributed variables with probability of success p, and let $x_i = 1$ denote the event where x_i fails. Then showing that $\frac{\bar{p}-p}{1-p}$ decreases in p is equivalent to showing that $\frac{p(x_1=1|S_n \leq k)}{p(x_1=1)}$ increases in p, where $S_n = x_1 + x_2 + ... + x_n$.*

Proof. First note that proving that the LHS of (??) is decreasing in p is equivalent to proving that $1 - \frac{\bar{p}-p}{1-p} = \frac{1-\bar{p}}{1-p}$ is increasing in p. Second, note that $1 - \bar{p}$ captures the expected proportion of failures of a Binomially-distributed variable given that the proportion of successes was at least k, or, put differently, given that the proportion of failures was at most k. Finally, we can replace the expected proportion of failures with the probability of a failure. All in all, we get that proving that $\frac{1-\bar{p}}{1-p}$ is increasing in p is equivalent to showing that $\frac{p(x_1=1|S_n \leq k)}{p(x_1=1)}$ is increasing in p. $\square$

Proving the proposition

Proof. Lemma ?? implies that if $x_i = 1$ denotes the event where x_i fails, and S_n counts the number of failures among n trials, then we need to show that $\frac{p(x_1=1|S_n \leq k)}{p(x_1=1)}$ increases in the probability of success p. For tractability, we will now revert to the more standard notation of $x_i = 1$ as denoting the event where x_i *succeeds* (s.t. S_n counts the number of successes among n trials), and show that $\frac{p(x_1=1|S_n \leq k)}{p(x_1=1)}$ *decreases* rather than increases in the probability of success p.

Version 1

$$\frac{p(x_1 = 1|S_n \leq k)}{p(x_1 = 1)} = \frac{p(s_n \leq k|x_1 = 1)}{p(s_n \leq k)} = \frac{p(s_{n-1} \leq k - 1)}{qp(s_{n-1} \leq k) + pp(s_{n-1} \leq k - 1)}$$

16

$$= \frac{p(s_{n-1} \le k - 1)}{q(p(s_{n-1} \le k) - p(s_{n-1} \le k - 1)) + p(s_{n-1} \le k - 1)} = \frac{p(s_{n-1} \le k - 1)}{qp(s_{n-1} = k) + p(s_{n-1} \le k - 1)}.$$

Version 2

$$\frac{p(x_1 = 1 | S_n \le k)}{p(x_1 = 1)} = \frac{p(s_n \le k | x_1 = 1)}{p(s_n \le k)} = \frac{p(s_{n-1} \le k - 1)}{qp(s_{n-1} \le k) + pp(s_{n-1} \le k - 1)}$$

$$= \frac{p(s_{n-1} \le k - 1)}{q(p(s_{n-1} \le k) - p(s_{n-1} \le k - 1)) + p(s_{n-1} \le k - 1)} = \frac{p(s_{n-1} \le k - 1)}{qp(s_{n-1} = k) + p(s_{n-1} \le k - 1)}.$$

Using a known result,[4] according to which

$$p(s_n \le k) = \frac{n!}{(n - k - 1)!k!} \int_0^q t^{n-k-1} (1-t)^k dt = \frac{n!}{(n - k - 1)!k!} \int_0^1 q^{n-k} s^{n-k-1} (1-qs)^k ds,$$

we get that $\frac{p(x_1 = 1 | S_n \le k)}{p(x_1 = 1)}$ is decreasing in p if and only if $\dfrac{q \binom{n-1}{k} p^k q^{n-k-1}}{\frac{(n-1)!}{(n-k-1)!(k-1)!} \int_0^1 q^{n-k} s^{n-k-1} (1-qs)^{k-1} ds}$ is decreasing in q, i.e., if and only if $\dfrac{\binom{n-1}{k} p^k}{\frac{(n-1)!}{(n-k-1)!(k-1)!} \int_0^1 s^{n-k-1} (1-qs)^{k-1} ds}$ is decreasing in q. Thus, it is sufficient to show that $\int_0^1 s^{n-k-1} (\frac{1-qs}{1-q})^{k-1} \frac{1}{1-q} ds$ is increasing in q. Since $(\frac{1-qs}{1-q})^{k-1} \frac{1}{1-q}$ is non-decreasing in q for every $s \in [0, 1]$, the proof is complete.

$\square$

---

[4] See equations (3) and (4) in https://mathworld.wolfram.com/BinomialDistribution.html.