Alliance Games and Racism

by
Malcolm Murray
University of Prince Edward Island

Abstract of "Alliance Games and Racism" by Malcolm Murray

This paper reports on recent game-theoretic results which through some light on the regular recurrence of racism in a wide variety of societies. It is shown that racism is a niche strategy which can arise and be stable in a variety of circumstances.

Papers from the 2003 meeting of the Canadian Association for Reductionist Philosophy
"Alliance Games and Racism" by Malcolm Murray, University of Prince Edward Island

Racism strikes us as irrational, not merely immoral. My defecting against one group (so long as it is not unilateral) means merely that I fail to reap the benefits of mutual cooperation. Yet despite our best efforts, racism is still rampant across societies. A descriptive explanation for the lingering success of racism is that racism has evolutionary fit. This does not mean that racists do "best" under all conditions. Simply (and perhaps horribly enough) there will be niches for racist strategies under certain conditions. Nature prefers polymorphisms, however, and so our abhorrence of racism also finds a place.

The success of operationally defined moral agents in PD games, Ultimatum games, and games of Chicken are due in large part to strategies that correlate with one another. If like plays like, those who are fairminded will do better than those who are less fairminded. Benefits to cheaters accrue only in cases where cheaters are allowed into a fairminded population. Thus, the success of fairminded strategies is contingent on their finding a way to keep the different out.

This wonderfully successful and simple moral strategy has an ugly side-effect. Sameness in terms of dispositions to cooperate can too easily be associated with irrelevant features, like skin color, sex, or cultural differences in gestures, facial features, or conceptions of social space. If such differences cloud our ability to detect propensity to cooperate, the prudent resolve is to defect. Heuristic shortcuts can therefore develop in which irrelevant differences become markers for non-correlation. In other words, racism, classism, and sexism may be the progeny of moral strategies, not immoral strategies. It is not surprising in biology that similar structures perform contrary tasks.

A Note about Evolution and Ethics

In this paper, I track evolutionary fitness, not rationality. Some defense, perhaps, is needed for this shift. After all, if a thing's survival can be explained by its having evolutionary fit, and its evolutionary fit is defended on the basis of its survival, we have a vacuous theory. Moreover, how could a descriptive account of evolutionary fitness provide any insight into the normativity of ethics? Nothing, say many, including such notables in evolutionary models as John Maynard Smith: "A scientific theory – Darwinism or any other – has nothing to say about the value of a human being."[1] Evolution seems to show that groups are differentiated and some are destined for extinction. To infer from this that therefore it is *right* that they be so extinguished (as the Social Darwinist movement professed) is to unfoundedly assume that what *is* is also what *ought* to be. So the argument that evolutionary theory cannot speak toward ethics is to deny that description has anything to do with normativity.

One needn't deny that *is* cannot by itself imply *ought* to suggest that evolutionary models can have something to tell us about ethics. The problem with Social Darwinism, for example, is not (merely) that they violated the is-ought distinction: they got their science wrong as well. For one, they fail to recognize that evolution, in David Hull's terms, "is the process by which rare alleles become common, possibly universal, and universally distributed alleles become totally eliminated."[2] A duty to remove that which doesn't fit the norm presupposes there is a norm. More than that, the Social Darwinists failed to consider our ethical beliefs as themselves being part of our evolution. To say we ought to excise the sick and infirm since they do not contribute to good offspring, is to miss that we tend to find such sentiment morally repugnant. From where did this repugnance originate? It strikes me that a larger account of evolutionary dynamics should have at least *something* to say

1. John Maynard Smith, "Science and Myth," in David Hull and Michael Ruse (eds.) *The Philosophy of Biology* (Oxford: Oxford University Press, 1998) 374.
2. David Hull, "On Human Nature," in David Hull and Michael Ruse (eds.) *The Philosophy of Biology* (Oxford: Oxford University Press, 1998) 392.

about this. Evolutionary fit is the right ordering of the right amount of a vast number of characteristics and dispositions – including social behaviors. If inherent value is something we ascribe to human beings, then it is not improbable that that ascription has evolutionary fit.[3] To speak of the *ascription of value* to human beings is to speak about one of the features (our propensity to ascribe value to human beings) that may be part of the Homo Sapiens arsenal for survival. If so, one would think evolutionary theory can account for it. After all, if the moral values we routinely ascribe to persons has no connection to the evolutionary stability of Homo Sapiens, this would count as a counterexample to the so-called vacuous principle of evolutionary theory – that everything which survives must have evolutionary fit. If moral behavior really disadvantages owners of that disposition in terms of evolutionary fit, we should predict the weeding out of morality. Conversely, if morality has remained, we should be able to track its evolutionary benefits.

The findings of Axelrod, Binmore, Skyrms, Gauthier, and Danielson all show how moral development *pays* the moral agent, despite the seeming paradox that morality is to curtail (some of) our purely self-interested actions. Generally speaking, agents exhibiting operationally defined moral dispositions do better than agents lacking moral dispositions. A closer look is not quite so rosy. Danielson's (seemingly) less moral agents fare better than Gauthier's moral agents. Skyrms offers support to a Rawlsian picture of justice as fairness in his *Evolution of the Social Contract*. Dispositions prone to offer only fair distributions and to accept only fair distributions have evolutionary fit. He fails to emphasize, however, under what narrow conditions this result ensues. Starting with equal population proportions will have more egoist strategies win. Moreover, altering Skyrms' initial population distribution of egoists and fairmen by even .001 will be sufficient to tip the scales back in favor of egoists.[4] Skyrms has certainly identified a condition in which fair minded persons will have evolutionary fit, but he has also shown (although he did not emphasize this point) that the odds are not in its favor.

All that we can so far say, then, is that moral-*like* behavior has evolutionary fit, or that moral behavior has evolutionary fit in exceedingly rare cases. These are not overwhelmingly favorable results for tracking the evolution of moral behavior. The "successful" disposition does not perfectly match the going account of morality. On the basis of this gap, we could conclude *either* that evolutionary models have nothing to teach us about morality, *or* that our ideals of morality are poor approximations or over-generalizations of the merits of morality. Taking the evolutionary model's tautology seriously commits us to the latter.

Since we have attitudes against racism, this should be accounted for on an evolutionary model. At the same time, racism is rampant throughout human history, and this *too* needs to be accounted for on an evolutionary model. Typically racism is viewed as an aberration of moral

---

3. Admittedly, not all features of a given species necessarily contributes to that species's survival. Hair growing out of one's ears in old age does not obviously contribute to the survival of Homo Sapiens, for example. Since nature generally abhors unnecessary appendages, however, the odds favour the attribution of morality playing a contributing role for our kind.
4. There are eight strategies in Skyrms's Ultimatum Game: S1-S8. S1 is called Gamesman and S7 is called Fairman. For the details of the game, see Appendix 1. Skyrms' initial population proportions in the case where S7 wins is the following: S1 = .32, S2 = .02, S3 = .1, S4 = .02, S5 = .1, S6 = .02, S7 = .40, S8 = .02. With these initial population proportions, S7 ends up with 56% of the total population along with another moral-like strategy S5 at 44%. Why S7 would be so favoured at an initial population is questionable. Ignoring that, if we alter S1 to .321 and S7 to .399, after 78 generations, S1 (Gamesman) completely takes over the population. Skyrms has certainly identified a condition in which fair minded persons will have evolutionary fit, but he has also shown that the odds are not in its favour.

behavior, or an example of pre-moral behavior. An account more consistent with evolutionary models is to say that racism is in fact an *output* of moral behavior.

A single strategy that produces wildly divergent behavioral patterns is not odd in biology. Despite the vast array of diversity found in nature, nature abhors complexity in terms of its basic building blocks. There is no reason to think strategies concerning social interaction are different in this respect.

A Note on Rationality and Heuristics

Rational strategies do not necessarily prevail in nature. Evolution tends to favor simpler – often patently irrational – strategies. A danger of appealing to simple strategies in social interactions is that they tend to be less sensitive to their external environment. Successful agents require adaptability. Theories of rationality provide an ideal model for context sensitive agents. Evolutionary mechanics seem desirous to ignore or override rationality. A tension exists, then, between simplistic algorithms and the need for context sensitivity provided by ideal rationality.

By ideal, I mean there may be a gap between what a person does and what they should have done rationally speaking. It is because rationality is an ideal, that we can say of someone that he behaved irrationally. Gaps need not be bridged. That rationality is an ideal does not imply that we ought to follow it any more than we ought to follow the moral ideal of turning the other cheek when struck.[5]

Since rational theories vary just as moral theories vary, saying I have no need to follow what is rational may indicate merely that I'm using the wrong model of rationality. Let us consider the generally received view that a model of rationality entails maximizing one's personal utility. On such an account, to entertain the possibility that I have no need to follow the advice of rationality reveals incoherence. Embedded in the utility maximization model lies the concept of personal utility. Acting rationally will secure me more of what I prefer. So to say I might not be interested in following the advice of rationality is plainly to confess that I do not prefer that which I prefer: an incoherence. Imagine being presented with a choice between two free lotteries: A and B. A offers a 50% chance of winning $50. B offers a 20% chance of winning $100. Even people who prefer more money to less, other things being equal, tend to choose B over A. It takes some theory of rationality to tell us that choosing B is inconsistent with the claim that one prefers more money to less.

I accept all this. Still, at the biological level, being ideally rational is itself something that has to be shown to have evolutionary fit.[6] To be rational requires brain power. To have sufficient brain power requires having the requisite brains. To have the requisite brains requires having the requisite bodies to maintain it. To have the requisite bodies, requires having the requisite resources. These are scarce, and in any event fluctuate.

Admittedly, if we can theorize about an ideal rational choice, that presupposes our brains are already sufficiently developed for ideal rationality. Still, being moved by heuristics may be quicker. A Lamarckian picture of evolution looks to what is needed (for example, the ability to reach branches higher than competitors, or the ability to follow ideal rationality) and then sexually selects those members of the species most having those desired features. So giraffes grow longer necks. That we don't typically notice women sexually selecting ideally rational nerds over less rational hunks is all the evidence required to defeat a Lamarckian model. The Darwinian theory has it rather different.

5. A danger in rejecting rationality *as a whole* can be observed by asking on what basis one is rejecting rationality. If rationality itself, the argument may be rejected by its own conclusion. If by an appeal to irrationality, that too will have no hold on defenders of rationality.

6. Nozick provides an account of this. Robert Nozick, *The Nature of Rationality* (Princeton: Princeton University Press, 1993) 107-132.

An aberration of long neckedness will benefit from its ability to reach leaves on trees untouched by competitors. Greater dietary source means greater chance of survival. A greater chance of survival means a greater chance of offspring for that mutant. So long as reproduction continues to favor the long-necked feature, the long-necked giraffe slowly takes over the population of giraffes. But there is always a trade off. A hugely long neck may be as disadvantageous as a short neck. Likewise, rationally-*like* individuals may evolve, while ideal rational agents may not. So long as certain heuristics can develop which work in normally occurring situations, the speed and convenience of these cheap heuristics may benefit agents who have developed them more than agents who have not. The cost of ideal rational models may not be sufficiently compensatory, given the environments they generally inhabit.

Admittedly there are many cases in nature where heuristics backfire. The program in a greylag goose to retrieve stray eggs to her nest can get fooled by footballs, skulls, and lightbulbs. The drive for male red-winged blackbirds to defend their territory is triggered not merely by the red wings of invading male red-winged blackbirds, but red balls and red shirts of passing joggers. The sexual attraction to bigger claws in Fiddler crabs results in offspring that are both more attractive to prey and less able to avoid prey. Vibrations on a web trigger eating behaviors in female spiders that need to be circumvented by amorous male spiders (by repeated tapping on the web until the eating response is saturated). The success of a hawk in a world of doves is not carried over in a world of Hawks. In cases of humans, anything involving probability reasoning is a prime example of heuristics coming up short. But, so long as the probability of agents finding themselves in such contexts is low, the rewards of heuristics may outweigh the benefits of adopting decision making on the basis of ideal rationality.

This is the same with morality. We may well end up with a distinction between ideal morality and a heuristic morality tracked by evolutionary models, but nothing commits us to move toward the ideal morality merely by pointing out this gap.

A caveat is needed. Since environments are variable in both the short term and long term, plasticity is necessary if organisms are to survive to reproduce. The danger of relying too much on heuristics is to rely too much on the chances that the environment in which the heuristic works will remain. If biology teaches us anything, that assumption is not to be expected. In fact, precisely because environments are not stable, variability is essential to biological species. This does not mean we will abandon heuristics. It is not a matter of choosing a context-insensitive heuristic or an ideal rationality. Heuristical programs can be fine tuned, and some may be more sensitive to environmental fluctuations. Successful heuristics will be those more sensitive to the varying conditions. Let us say that to be fully sensitive is to be rational in the ideal sense. But sensitivity to one's environment can be on a continuum.

Modeling Racism: The Alliance Game

Why would racism have developed? It would seem prudent to resist cooperating with someone in whom you have no trust. Trust is often based on behavioral cues, posture, and facial expressions. Pre-verbal communication between races is more opaque than within races. If we cannot read the disposition type of our playing partner, the prudent resolve is to defect. If we define racism as a tendency to defect against members of other races, racism is seen to be fairly predictable. It may well be mere opacity between races that explains our racist tendencies.

Psychological studies supporting such tendencies abound. Real persons playing prisoner dilemma games in experimental situations will vary their moves according to how the other player is identified. They are more likely to offer initial cooperative moves if the other player is identified as an in-group (a teammate, for example), and more likely to offer initial defect moves if the other player is identified as a member of an out-group (someone from a rival university, for example).[7]

What is important is that groups distinguish themselves from others, and not that particular groups exist. That is, if it weren't Jews that the Nazis hated, another group would have done just as well. For any group, a further distinction can be made which can be used as markers for racism. The forming of alliances, then, is all we need to track the dynamics of racism.

My findings are consistent whether we play Ultimatum Games or PDs. To track racism is to take whatever strategies there are in a given game and further divide those strategies into three sub-groups: Racists, Non-Racists, and Racist Targets. In the Ultimatum Game, there are eight strategies: therefore, to track racism in the Ultimatum Game, we require twenty four separate strategies. Beginning with four strategies in the PD, we would need to track twelve strategies to model racism. In both cases, we can do with slightly less dispositions, but we can simplify matters enormously by focusing on what I shall call the Alliance Game. That is, the results from the Alliance Game are similar to the results in the Racist Ultimatum Game (see Appendix 1) and the Racist PD (see Appendix 2).

In a world initially populated by Unconditional Cooperators (UC) and Unconditional Defectors (UD), and Conditional Cooperators (CC), CC agents can completely take over the population in 1026 generations.[8] Varying the initial population proportions can provide some small niches (less than 1%) to UCs, but in no case can UDs survive. Let us then assume the success of CC. (To see the results with Danielson's Reciprocal Cooperators in the mix, see Appendix 2.) In such a world, everyone's benefit means no one's relative gain. When relative gain matters, there will be incentive to tip the scales. Crude attempts to try to exploit one group will fail. Since each group in this population is comprised of conditional agents, Group A's attempt at exploiting B will mean merely a smaller payoff for A (as well as B). Group C, if such exists, will in fact benefit from A's greed. If A and B fight each other while each continuing to cooperate with C, C will be the winner of this population.

---

7. Eg. Peter Kollock, "Transforming Social Dilemmas," in Peter Danielson (ed.) *Modelling Rationality, Morality, and Evolution* (Oxford: Oxford University Press, 1998) 185-209.

8. What ought we to mean by a "conditional" agent? If we define a conditional agent as one who bases her strategy on the dispositions of the other players, then many of the successful strategies in the computer simulation trials of Axelrod and Skyrms are unconditional agents. If conditional strategies require rational deliberation, an appeal to the evolution of morality will fail. We do not want to say that conditionality is that which cannot be instantiated in a robot or computer program, for Gauthier's and Danielson's models can be so instantiated. In other words, we don't wish to conflate conditionality with the concept of free will or human agency. Axelrod's (or Rapoport's) Tit-for-Tat (TFT) agent is not conditional in one sense since it merely mimics the previous move of its co-player. But TFT is a conditional agent in the sense that its cooperation is conditional upon the last move of its co-player. In the Ultimatum Game, Skyrms's successful Fairman offers an equal distribution of manna, and accepts only an equal distribution of manna. If anyone offers him less than what's "fair," Fairman will reject it. Whether we interpret this as a conditional move or not will depend upon what we think conditionality must be predicated. In any event, Skyrms' emphasis on correlation clearly matches the conditionality present in Gauthier's and Danielson's models. Fairmen will only play fair on their first move, but in the correlated conventions, Fairmen first selects who she will play with. In this case, it will be on the basis of whether the other player is a Fairman or not. In this sense, we can say, Fairman will play fair on the first move *on the condition that* the other player will also play fair on the first move. Moreover, the success of cooperators over defectors can be shown *without* the introduction of Gauthier's and Danielson's conditionality. So long as players may correlate on the basis of space or markers for sameness (like skin colour?) cooperators can survive alongside defectors. So long as the correlation is greater than 50%, cooperators will in fact take over the population.

The existence of group C, however, will give A an alternative strategy. A may say to C: "Look, we'll continue to cooperate with you only on the condition you defect against B." Even if successfully convincing C to defect against B, this strategy still won't help A – neither relative to C, nor relative to A's own prior utility. Beyond that worry, C may be motivated to retaliate against A. An exploitive neighbor is not always going to work in C's behalf. Given that conditional strategies predominate in each group (in our scenario), if C now defects against A and cooperates with B, B and C will do equally well, and A will suffer relative to B and C. Not only will A fail to achieve top status relative to the other groups, it runs the risk of falling to the bottom rung.

It is premature to conclude that racism fails. To do so assumes the population proportions are equal among A, B, and C. A's reasoning is sound if in terms of population proportions, A is larger than C, and C is larger than B. In other words, the majority exploiting the minority can pay.

Assuming A offers C a real choice, we will need to divide C into two classes, C1 who accepts A's offer, and C2 who doesn't. The proportion of C that is C1 can vary between 0 and 1. A will cooperate with A, will defect against B, will cooperate with C1, and defect against C2. B will defect against A, cooperate with B, defect against C1, and cooperate with C2. In terms of C, a choice point needs to be made concerning the interaction between C1 and C2. Will they cooperate or not? If they do not, then C1 will simply be an A. To avoid this conflation, let us then imagine Cs cooperate among themselves irrespective of their alliances with A.

Let us assume the following utilities: Defect,Coop = 2,-1. Coop,Coop = 1,1. Defect,Defect = 0,0. Coop,Defect = -1,2. The payoffs for our agents, ignoring population proportions, then can be seen in the following table:

	A	B	C1	C2	Average
A	1	0	1	0	0.5
B	0	1	0	1	0.5
C1	1	0	1	1	0.75
C2	0	1	1	1	0.75

Here we assume that B is a conditional agent and can successfully thwart A's and C1's exploitation by defecting in retaliation. Assuming this, the C is the benefactor of A's greed, and this is so whether C accepts A's offer, or rejects A's offer.

But something different occurs once we take population proportions and replicator dynamics into effect. (For an account of the replicator dynamics, see Appendix 3.) So long as A outnumbers C who in turn outnumbers B, A may stand to benefit from its seemingly counter-productive discrimination.

Let us assume the proportion of A is .60, C1 is .15, C2 is .15, and B is .10. There is nothing magical in this division. We can imagine a world where A represented the bulk of CCs, and over time cosmetic variations on CC developed, such that B and C represent offshoot mutations. This could account for the low proportion of their existence in the total population. Nothing hinges on this, except to show that there exist conditions in which one group of CCs may come to prosper from exploiting another group of CCs.

In one generation, A and A's ally, C1, grow in size. B and its ally, C2, dwindle. In only three generations, the minority group B is extinguished. In six generations, an equilibrium is reached where B and its ally C2 are extinct. A dominates the population at 74%. C1 has 26%.

By the racist actions of A, B is extinguished and A dominates. Racism has evolutionary fit. In cases where A is only slightly larger than B, such results do not occur. Assuming the following initial populations: A = .35, B = .32, C1 = .165, and C2 = .165, an equilibrium is reached after fourteen generations where C1 dominates at 73%. The population is comprised of 23% of C2,, and As comprise only 2% of the population. C2's success piggy-tails on C1's success. Still, B is eliminated from the genetic pool. This shows how vulnerable minority groups are to the whims of others. Can a portion of the racist A group now target C2? We have already witnessed the fact that they can't do so *directly*. If A is now divided into A1 who continue to cooperate with C2, and A2, who defects against C2, A1 will prosper relative to A2. The A2 mutation cannot thrive. A2 needs a third party to threaten in order to be successful with its racist interactions against C2. If none are forthcoming, the initial success of racism will nevertheless lead to an equilibrium of non-racist players.

So long as A outnumbers both B and C2, A and C1 take over the population. If A has less than 50% of the total population, their racist activities drive them to extinction, *unless* C1 outnumbers C2 and B consumes less of the initial population proportion than does A. If equal population proportions exist across the four groups, C1 and C2 take over the population. In these conditions, A's greed annihilates B, but A as well. It is a Pyrrhic victory. Interestingly, whether one forms an alliance with A or not, will not matter. (See the varying outcomes from shifts in initial population proportions in Appendix Three.)

Conclusion

In the Alliance Game, the full Gauthier-Danielson PD Game, and Skyrms's Ultimatum Game, there is a niche for racism. Admittedly not all initial population proportions will favor it any more than all initial population proportions favor Fairmen or CC. Racist strategies are simpler strategies than conditional cooperation, since the marker that triggers cooperation is visual appearance. The marker that indicates cooperation for conditional cooperators is far more subtle: more open to error and exploitation. If we build in the costs of these errors for non-racist conditional cooperators, racist strategies fare even better.

In all but a few of the computer simulations, the targets of racism are eliminated. It is an odd victory to say that racism will be eliminated on the basis that the victims of racism will become extinct. So although the findings of my computer trials may explain why racism is so rampant (when initial conditions favor it and given the ease of coordinating on the basis of visual markers), it fails to explain why the target groups of racists continue to survive in our world.

Still, the results point toward understanding racism in a different way than typically reported. Understanding that the root of both racist and non-racist behaviors is the same basic conditional strategy produces a new (perhaps disturbing) conception of moral behavior.

APPENDICES

Appendix 1: Racism in The Ultimatum Game

For the Ultimatum Game, imagine a pie divided into ten pieces. One person offers a division, and the other accepts or rejects the offered division. If the second rejects the division, no one gets any of the pie. The fear of getting no pie rather than some pie would make (one would think) offerers wary of seeming too greedy. Certainly an offer of "10 for me and 0 for you," stands to be rejected, leaving 0 for them both.[9] But is it rational for the hungry to reject an offer of one piece of pie? If not, it would be rational to offer "9 for me, 1 for you," predicting acceptance.

Thankfully Skyrms limits the range of offers to two: a 9-1 split and a 5-5 split. With two possible first moves (Demand 9 or Demand 5) and four possible second moves (Accept all, Reject all, Accept 9 but Reject 5, Reject 9 but Accept 5),[10] we are left with eight strategies: (1) Demand 9; Accept All. (2) Demand 9; Reject All. (3) Demand 9; Accept 5, Reject 9. (4) Demand 9; Accept 9, Reject 5. (5) Demand 5; Accept All. (6) Demand 5; Reject All. (7) Demand 5; Accept 5, Reject 9. (8) Demand 5; Accept 9, Reject 5.

Those who Demand 9 will do well against those who Accept 9. We can think in terms of their getting 9 pieces of pie rather than 1, where more pie is better (no diminishing return; no cholesterol worries). Skyrms is not testing "rationality," though, he is testing evolutionary fitness. How well one does depends on who else is in the mix (or who has died off, and what new agents invade). Thus "doing well" for Skyrms is translated as passing on one's genes. In this case, increasing the number of offspring for that player type. Thus, the next generation increases by (say) 9 "Demand 9ers." If the "Accept 9s" fare poorly, there will be fewer of them relative to the increase in the more successful strategies. Each generation then competes in different population mixes and these differences in population mixes will impact differently on the success of any given agent. For example, #1 (Demand 9; Accept all), will do better playing with #5 (Demand 5; Accept all) than playing with one of her own siblings (14 slices vs 10). #1 will do worst playing either #2 (Demand 9; Reject all) or #3 (Demand 9; Reject 9, Accept 5).

Skyrms ran 10,000 generations of this game (where every player plays every disposition per generation). Nicer players did not end up too badly so long as the initial mix favored them. So long as we start with a 30% population majority of #7 (Demand 5; Accept 5, Reject 9 (or "Fairmen" as Skyrms calls them)), immoral or unfair agents (those who Demand 9), die off. We are left with the completely obsequious #5 (Demand 5; Accept all (whom Skyrms called EasyRider)) (36%) and Fairmen (64%). Starting with an equal distribution of all types, however, yields less morally favorable results #1 (whom Skyrms called Gamesman) (87%) and #4 (whom Skyrms called MadDog) (13%).) Why should we anticipate favorable initial mixes? Perhaps "fluke" is as sufficient an answer as we need. After all, that's an acceptable answer in general talk of evolution. We do not require a reason for a mutation.[11] The important point for now is that Skyrms has tracked the conditions under

---

9. Notice how this counteracts the concept of Pareto improvements. Paretianism would demand *Accept*, since no one is worse off and at least one person is better off. But claiming that it is rational, let alone morally obligatory, to have to accept an offer of 0 of a social pie merely because the offerer wants it all for himself seems insane. (If it were a *private* pie, that would be a different matter.)

10. "Accept 9" stands for accepting the demand that the other gets 9 while you get 1. That is, "Accept 9" really means "accept 1 lowly piece of pie."

11. The argument for polymorphism can be made stronger. Skyrms's results are predicated on assuming offspring are clones of parents. Biological evolution permits random mutations or at least recombinations. So if "Demand 9; Accept all," for example, does better, it is not the case that all of its children will be "Demand 9; Accept all" clones. Some might be "Demand 9; Reject all." Incorporating a proportion of random mutations in the replicator dynamics will lower the odds of an equality condition.

which evolutionary dynamics carries the various strategies to a polymorphism that includes operationally defined moral agents who are, nevertheless, "weakly dominated, modular irrational strategies."

To model racism in this game, we need to further divide Skyrms's strategies into two groups who differ from each other only in terms of skin color: Let's call one group W and one group B. Let us next divide the Ws into two groups: racists (whom we'll designate as R), and Non-Racists, (whom we'll designate simply as W). R's will treat Ws differently than Bs. Rs will treat Ws according to whether the R is a Gamesman or a Fairman, etc.. When dealing with Bs, however, all Rs (no matter their dispositions when dealing with Ws) will demand 9 when moving first, and Reject all when moving second. Bs, meanwhile, will follow merely the heuristics of their given strategy (varying from S1 to S8).

Following Skyrms initial population proportions that favored Fairmen over Gamesmen, we arrive at an equilibrium of 36% EasyRiders and 64% Fairmen.[12] If we also assume an equal population distribution among Rs, Ws, and Bs, a polymorphic equilibrium is reached after 3019 generations in which non-racists beat out their racist counterparts. In fact, Non-racist Fairmen consume nearly half of the population. (Non-Racist Fairmen: .49, Non-Racist EasyRider: 0.27, Racist Fairmen: 0.16, Racist EasyRider: 0.09).

In the case where 50% of the population are Non-Racist whites, 20% are White Racists, and the remaining 30% are the minority B group targeted by racists, racists (as well as the minority Bs) are extinct. After 21,000 generations, White Non-Racist Fairmen consume 63% of the population, and White Non-Racist EasyRiders make up 36% of the population.

When the initial population has racists consuming 50% of the population, racists will keep their top status. The following equilibrium is reached after 966 generations: White Racist Fairmen: .38, Racist EasyRiders: 0.22, White Non-Racist Fairmen: 0.26, White Non-Racist EasyRider: 0.15.

When the initial population favors the non-racist minority group (50%), the Racists at 30%, and the non-racist Whites (20%), only then can some of the minority groups continue to survive. After 4538 generations, an equilibrium is reached in which White Non-Racist Fairmen = 0.49, White Non-Racist EasyRiders = 0.27, Black non-Racist Fairmen = 0.14, and Black non-Racist EasyRiders = 0.1

In each case, the fact that non-racists do better than racists is promising for those of us who maintain that racism is immoral. Two problems taint this happy picture. First, the existence of racists clearly impair the existence of the racist target groups, even in cases where racists fare poorly. Secondly, this tracks only the dispositions of Fairmen and EasyRider. When examining the effects of racism on all logical dispositions in the Ultimatum Game – even when original dispositions of strategies favors Fairmen and EasyRiders, Gamesmen take over the population. 64% of these Gamesmen are non-racist, and 36% are racist. That non-racists prevail over racists in this instance is hardly satisfying, since neither group is operationally defined as moral.

In the case where non-Bs outnumber Bs 3:1, and the non-R outnumber R 3:1, and the initial population favors Fairmen (70%) while Gamesmen have only 5%, the racist strategy is eliminated. After 2000 generations, White Non-Racist Fairmen = .80, White non-Racist EasyRider = .14, Black Non-Racist Fairman = .05, and Black Non-Racist EasyRider = .01.

So long as the initial population proportions favor non-racists and there is a smaller initial target populations for Racists to exploit, racism will become extinct. Otherwise, racists have a niche in the Ultimatum Game.

12. See Skyrms, 31.

Appendix 2: Racist Players in PDs

Let us consider a PD tournament with four dispositions, Unconditional Cooperators (UC), Unconditional Defectors (UD), Conditional Cooperators (CC), and Reciprocal Cooperators (RC). UCs cooperate always. UDs never cooperate. CCs cooperate with anyone else prone to cooperate. RCs cooperate only with those cooperators whose cooperation is conditional. Thus, while CC will cooperate with UC, RC won't. If the populations are equally divided among the four strategies, an equilibrium is reached in which RC has 58% of the population, and CC 42%. CC can dominate so long as the initial population of RC is lower than CC. For example, if the initial populations are UC: 0.3, UD, 0.5, CC, 0.2, RC 0.1, an equilibrium will be reached in which CC makes up 55% of the population, and RC 45%.

To introduce racism in this game is to divide the four dispositions into three sub-categories: A racist group, a non-racist control group, and a racist target group. We can designate the first group as R, the second as W, and the third as B. We can simplify our number of strategies by recognizing that unconditional agents cannot be racist. UD defects against everyone anyway, so will not distinguish between B and W. A UC racist, if such were possible, would do better than his UC counterparts, but we shall assume the unconditional nature of UC will prevent his ability to defect against any Bs. Thus, we have ten separate strategies or dispositions. Of course the dispositions or strategies of unconditional Bs (UC(B) and UD(B)) will behave just as their W counterparts (UC(W) and UD(W)), so it might be argued that no new strategy exists, but their different race markers (W or B) will impact their scoring.

Two scoring procedures are possible in this game. Rs may exploit or exclude Bs. If they exploit Bs, Rs will be a very successful strategy. Bs need not be so guileless, however, at least not CC(B) and RC(B). If our conditional Bs recognize when they are playing with Rs, these Bs will do well to defect. In the results which follow, I shall assume the latter.

Assuming equal population proportions across the ten strategies, racist strategies do better than their non-racist counterparts. RC(R) = 33%, CC(R) = 28%, RC(W) = 22%, and CC(W) = 17%. The racist targets have become extinct. Without racist targets, Rs are equivalent in practice to Ws. Therefore, the four groups collapse into two: RCs consuming 55% and CCs 45%.

If the initial population proportion favors non-racists over racists, it will pay to remain non-racist. For example, if RC(R) begins with .07 of the population, compared with RC(W)'s .2 and CC(R) having .1 of the population compared to CC(W)'s having .25, the following equilibrium results: CC(W) = .36, RC(W) = .30, CC(R) = .2, and RC(R) = .14. Although non-racism is better in terms of replicator dynamics, the niche for racism still exists, and that niche spells disaster for the existence of any racist target group.

Appendix 3: Replicator Dynamics for the Alliance Game

Let a, b, c1, and c2 represent the population proportion of A, B, C1, and C2 respectively. The formula for individual scoring (using the table in the text) is as follows:

$A = 1(a+c1)+0(b+c2)$
$B = 0(a+c1)+1(b+c2)$
$C1 = 1(a+c1+c2)+0(b)$
$C2 = 1(b+c1+c2)+0(a)$

Notice this does not exclude self-play. So long as the raw numbers of players is sufficiently large, this omission should be negligible. Throwing in our initial population proportions (p(a)) favoring A, the following utilities (U(a)) emerge:

$A = 1(.6+.15) = .75$
$B = 1(.1+.15) = .25$
$C1 = 1(.6+.15+.15) = .9$

C2 = 1(.1+.15+.15) = .4

Assuming the utilities are to be cashed out in terms of replicator dynamics, we can map the population progress of each of these groups over a number of subsequent generations. We can chart the replicator dynamics for strategy *A* using the formula: U(*a*)p(*a*)/**U**. U(*a*) = the average utility for a player using strategy *A*. p(*A*) = the current population proportion of players using strategy *A*. **U** = the average utility of all players. Determining the population proportions for the subsequent generation of our four groups of conditional cooperators, then, proceeds as follows:

U = .75(.60)+.25(.10)+.90(.15)+.4(.15) = .67

Therefore, the population proportions of the subsequent generation is determined in the following manner:

U(a)p(a)/**U** = .75(.6)/.67 = .67
U(b)p(b)/**U** = .25(.1)/.67 = .04
U(c1)p(c1)/**U** = .90(.15)/.67 = .20
U(c2)p(c2)/**U** = .40(.15)/.67 = .09

Results. For the following population sets, recall that A is the racist strategy and C2 is the Anti-racist strategy. Note that A wins in population sets 1,2, and 5.

1. A @ .5 = .73
   B @ .25 = 0
   C1 @ .1 = .27
   C2 @ .15 = 0

2. A @ .5 = .70
   B @ .25 = 0
   C1 @ .15 = .30
   C2 @ .1 = 0

3. A @ .25 = 0
   B @ .25 = 0
   C1 @ .25 = .5
   C2 @ .25 = .5

4. A @ .4 = 0
   B @ .3 = 0
   C1 @ .1 = .5
   C2 @ .2 = .5

5. A @ .4 = .53
   B @ .3 = 0
   C1 @ .2 = .47
   C2 @ .1 = 0

6. A @ .25 = 0
   B @ .5 = .73
   C1 @ .15 = 0
   C2 @ .1 = .27

7. A @ .3 = .04
   B @ .2 = 0
   C1 @ .3 = .96
   C2 @ .2 = 0

8. A @ .3 = 0
   B @ .2 = 0
   C1 @ .2 = .5
   C2 @ .3 = .5