

A 'DEFENSE' OF RACISM

I. SOME ETHICAL HOUSEKEEPING

1. The Context-of-Dissemination Objection

Time was when one could share one's research - or, so as not to exclude philosophers, one's thinking about something - without having to mention, let alone defend, *why* she might be researching this rather than that, or thinking about this rather than something else. Those times are gone. And - miss them though we might - it's probably good that they are. For it's not just that most research and thinking is done at public expense. Nor even that the answers we come up with have autonomous effects independent of the ends for which they were sought in the first place. It's also that the very *asking* of a question has effects independent of the reasons we're asking it.

For example, neuro-scientist Simon LeVay, himself gay, compared the mass of the hypothalami of gay men and straight men. He did so - and published his conjecture that heterosexuality varies directly with hypothalamus mass - on the supposition that, if confirmed, these results would put to rest the homophobic insistence that sexual orientation is a matter of choice, and that gay men could therefore just *decide* to join 'the rest of the human race'. What LeVay did *not* anticipate was that, in an already homophobic society, the more 'natural' conclusion is *not*, as it is with left-handedness and right-handedness, that people are just born differently, but rather that, "See, I *told* ya there was somethin' wrong with them fags!"

A similar worry is currently doing 'the circuit' over Thornhill and Palmer's *Natural History of Rape*. Put aside whether Dawkins' Selfish Gene Hypothesis is true. (Or even what 'true' could mean in this context!) The naturalization of rape is bound to lead to a biological determinist version of the Defense of Necessity: "But your Honor, I was just doin' what's in my nature!" Nor is the fact that we don't allow such a defense for killing one's cuckold, or innumerable other crimes, a guarantee the argument won't produce acquittals, or mere-slap-on-the-wrist sentences. So the question that cannot but be asked is, even assuming a naturalized account of racism would give social engineers greater prediction and control than do more 'politically correct' accounts, would the naturalization of racism give *solace* to the racist? And if so, would the distastefulness and other minuses of that solace outweigh the pluses of that greater prediction and control?

These are, of course, empirical questions, about which reasonable people can disagree. But I think racism falls into a category for which, fortunately, a more *principled* answer is available. If by 'racism' were meant nothing more than the acknowledgment of the occasional usefulness of racial categories, then we'd *all* be racists. Nor would we have any call to apologize for it. But this is *not* what racism is. It involves, as ineliminably, a set of judgments about the relative *values* of the races. And, parasitic on these

judgments, a set of moral *directives*, e.g. “There *should* be no mixing of the races!”

But value judgments and moral directives are notoriously vulnerable to second-order analysis. That is, in the same way that a man can't be thinking about what he's *doing* when he's having sex without losing his erection, neither can a racist sustain his bigotry and, at the same time, say to himself, “Ah, I see. My hatred of blacks is just an nth-ordered reinforcement mechanism for a coalition strategy in a looting game!” By contrast, one *can* simultaneously understand that rape is about genetic perpetuation and persevere in the rape. So, though the jury's still out for me on the context of dissemination for the naturalization of rape, I'm convinced, or at least satisfied, that there's nothing to be lost, and arguably a great deal to be gained, by naturalizing racism, if naturalized it can be.

2. The Robert Merrick Objection

So far I've considered, and dismissed, the worry that the naturalization of racism could be used to *perpetuate* the practice rather than deconstruct it. But I've *not* considered the objection that would be raised, *mutatis mutandis*, by Robert Merrick, the so-called Elephant Man, who's reputed to have cried out from within the cage in which he was being displayed, “I am not an animal! I am a human being!” That is, Merrick's need to assert his humanity was parasitic on its having been a live question at the time whether he was or wasn't fully human. Similarly, then, the very posing of a ‘purely scientific’, or ‘purely philosophical’, question about racism raises the possibility that racism might, under some circumstance, be *defensible*. And it's *this* possibility, even if never actualized, which, so long as it remains undischarged, is deeply offensive to those who've suffered from racism.

But, once again, the objection misfires. For it equivocates on the notion of defensibility. Many of us, myself included, view the recent conquest and occupation of Iraq as indefensible. But only a pacifist believes that the very *practice* of making war is, in this same sense, indefensible. More to the point, however, is the equivocation between the moral and purely *explanatory* uses of terms like ‘defensibility’ and ‘justification’. Murder is by *definition* morally indefensible, since it just *means* wrongful killing. But is anyone fearful that an explanation of *why* people wrongfully kill each other is likely to *encourage* wrongful killing?!

Moreover, our passing moral judgments and then acting on them is *itself* a practice, is it not? One which is the object of concerted philosophical enquiry. So even if the naturalization of racism *were* hurtful, what's at issue is not the hurtfulness of racism, nor the hurtfulness of naturalizing it. What's at issue isn't hurtfulness at all! Eavesdrop on the following conversation:

“Hey, nigger, back of the bus!”

“Excuse me, sir, but perhaps you didn't realize that when you call me that it hurts my feelings.”

“Oh! I'm sorry. Please forgive me. And here, have my seat!”

What makes this conversation nonsensical is precisely what makes *this* one so: “Please, Mr. Bush. When you make war on my country you run the risk of damaging property. Perhaps even hurting someone.”

“Oh! What could I be thinking?! Rumsfeld, call the hold thing off, this instant!”

In other words, to suppose that the hurtfulness of a practice is any kind of argument against it is to presuppose participation in *another* practice, namely the practice of not hurting each other. But if we could

presuppose participation in *that* practice, war and racism would be unintelligible to us. We have very little trouble making sense of war. So, say I, let's get up to like speed in our understanding of racism. Good philosophy presupposes only what *needs* to be presupposed to make philosophical enquiry possible. That we eschew hurting each other is not among these enabling presuppositions.¹ Or at least it's on this supposition that I give myself leave to proceed to ...

II. SOME CONCEPTUAL HOUSECLEANING

1. The Vacuity of the Creationist/Darwinian Debate

I'm going to indulge for a moment in a little, well, madness. But bear with me. There *is* method to it. So first:

There's probably no theory so widely embraced - and yet at the same time so consistently misunderstood - as Darwin's theory of evolution by natural selection. Among these common misunderstandings is, of course, that it's a *teleological* theory: that the object, or at least upshot, of an evolutionary process is the *improvement* of the species. Nuff said. But an even more pervasive misunderstanding - though one less commented on - is that it's an *historical* thesis: that it's committed to making claims about how certain states of affairs, or at least certain *features* of certain states of affairs, came *about*.

But it's not. Not really. It's true that if one were a Darwinian and a betting man, he'd probably bet that his ancestors did indeed come down from the trees. But he *need* make no such bet. He could, for example, subscribe, as I do, to the *Five Minute Hypothesis*, according to which:

The world came into being a scant five minutes ago, just as it *was* five minutes ago, each and every thing in it having precisely the properties it *would* have had five minutes ago *pace* the Five Minute Hypothesis, save for any and all *historical* properties it would have had five minutes ago *pace* the Five Minute Hypothesis.

But one who subscribed to the Five Minute Hypothesis would remain a Darwinian nonetheless, provided only that she believes:

The way the world will be five minutes *hence* is a function of which things are such that, given the way the world will be for the *next* five minutes, they'll *survive* those next five minutes, and which things are such that, given the way the world will be for the next five minutes, they won't.

But who'd deny this?! And so who could *fail* to be a Darwinian?!

¹ I'm not entirely satisfied with this response. There's a huge literature, initiated by Habermas and since carried forward largely by feminists, trying to set out the requisite norms of dialogical deliberation. A common desideratum seems to be respect. Inoffensiveness, on the other hand, seems to be too strong a condition for what's often needed in these contexts.

Yes, que sera sera. True, but trite. So perhaps what's at issue between Creationists and Darwinians isn't what will be but the mechanism giving *rise* to what will be. Suppose, for example, a Creationist allows that, yes, things perish, but, being pathologically squeamish about sex, she won't allow that anything actually reproduces. Things do, however, replicate. How? By God making more of the kinds of things He likes, and less of the kinds of things He doesn't. How is God's replicating only the things He likes any the less a feature of the world in virtue of which things are as they are than is the fit between drier land and a horse's hoof and the *misfit* between drier land and the *proto-horse's* hoof?! So divine providence is no less an evolutionary process than so-called *natural* selection. Thus Creationist and Darwinian accounts of the origin of species, albeit at odds, are only *uninterestingly* at odds.

Here's another misconception. We're told that, no less than alleles, memes too are naturally selected for or against. So, for example, *modus ponens* has been naturally selected *for*, and the only reason affirming the consequent hasn't been naturally selected *against* is that the world is such that the frequency and direness of cases in which *p* is accompanied by *q*, but not vice versa, don't warrant the computational costs of distinguishing between the two patterns. Fair enough. But now consider the *God's-a-Mother-Hen Hypothesis*, according to which:

Every time we reason, "If *p* then *q*, *q*, so *p*!", when in fact *p* is false, God rearranges the world so that *p* is true.

Under such circumstances, on what grounds could it be claimed that affirming the consequence is nonetheless irrational? None, if rationality is to be a normative notion, and if those counseling what epistemic norms to adopt are concerned - as, according to Descartes, is God - with "the best system that could be devised for the preservation of the healthy man"!

The upshot of all this skepticism and God-talk is not to suggest that one story, Creationism, say, is as good as any other, say, Darwin's theory of evolution by natural selection. Rather it's to show that what *really* matters about the stories we tell isn't their correspondence to reality, whatever that might mean. What matters is the *logic* such stories must satisfy. You say it's a Kahnemann-Tversky glitch in his algorithm for fertility detection that accounts for Tom's pedophilia. I say the devil makes him do it. What makes your story better than mine is not its more parsimonious ontology. (Demons are far more parsimonious than complexes. That's why complexes are *called* complexes!) It's that the devil can make us do pretty much anything he likes, and he can like pretty much anything. So citing his druthers offers very little in the way of explanation. A smooth skin heuristic in one's fertility detection algorithms, by contrast, will misdirect one's sexual attention towards children but not towards post-menopausal women. And it's *that* that makes it a better explanation of pedophilia. But - and this is an important 'but' - were we looking to understand why Dick raped the *elderly* woman, "The devil made him do it!" is as good an answer as a malfunction in his fertility detector. After all, a malfunction in Dick's fertility detector will explain anything and everything other than his making advances towards post-pubertal pre-menopausal women!

2. Some Game Theoretic Apparatus

But enough of these meta-considerations. Let's get down to the real work, beginning with some apparatus-

compiling:

A Dennettian agent is a postulated repository of behavior that behooves us, given our computational constraints, to treat *as if* it had beliefs and desires, whether it *really* has them or not, in whatever sense of 'really' the realist about beliefs and desires might have in mind when she deploys this word 'really'. Thus Dennettian agency is a *dyadic* relation. Strictly speaking, some *x* is a Dennettian agent to *some y*. A human being is a Dennettian agent to herself. Human beings are Dennettian agents to each other. And though that thermostat on the wall is not a Dennettian agent to *most* of us, were Socrates brought here in a time machine, chances are it *would* be a Dennettian agent to him.

The logic to be satisfied by any story about the *interactive* adaptations of Dennettian agents is called *game theory*. Such interactivity, or 'games', come in three varieties: games of pure conflict, pure coordination problems, and mixed-motive games. According to the reduction I want to urge here, what we folk-psychologically experience as *morality* and *politics* are just intra-mental and extra-mental responses respectively to, and only to, otherwise sub-optimal payoffs in *mixed*-motive games. So if an action is such that no cooperative dividend is available - i.e. it's purely conflictual - or if accessing the cooperative dividend is motivationally *unproblematic* - i.e. it's just a matter of coordination - it's in (what David Gauthier has called) the *morally free zone*. So, for example, to say, as 'they' are wont to do, that "All's fair in love and war!", is *really* to say that in these particular practices the very *notion* of fairness is misplaced. It's to say that love and war are morally free zones.

Just as an aside, I'm not sure about love, but I *am* sure 'they' are wrong about war. War is *not* a game of pure conflict. For as much as I'd like to appropriate your hillside, not if all I'm going to be able to plant there is my tombstone. And, fortunately, likewise feel you with respect to keeping me *off* your hillside. So, as I say, what makes rules of engagement - and terms of *disengagement* - possible is that war is a *mixed*-motive game. What makes me less sure about love is the current controversy over Thornhill and Palmer's *Natural History of Rape*. But let's leave that for another day.

What can be postponed only momentarily, however, is the role pure coordination problems play in the resolution of mixed-motive impasses, and the role of (what I shall be calling) *extra-logicity* in the resolution of these pure coordination problems. As I say, to that topic I shall be returning momentarily, and in some detail.

3. The Vacuity of the Struggle Over Ownership of Moral Language

In the meantime, however, what follows from this analysis of morality - and this has proven a source of both considerable embarrassment and even more considerable rancor - is that a mother's love for her child, insofar as it's not mixed-motive, is not a moral disposition. By contrast, choosing to tie your child's laces at the playground as well as mine *is* a moral matter, because much as I'd like to tie my child's but not yours when it's *my* day to supervise the playground, I don't want you tying your child's but not mine's when it's *yours*. Of course to access the cooperative dividend I'll have to convince you I'll follow *through* on our agreement to tie each other's child's laces. And to do that I may have to engineer in myself, and make transparent to you, a love for your child akin to the love I bear my own. But ...

But wait! If I bring myself to love your child, tying her laces will no longer be mixed-motivated! Worse yet, loving your child as a means is phenomenologically indistinguishable from loving your child

intrinsically. So, quite aside from Gilligan-esque concerns about defining most of what counts as female morality (e.g. that of mother-love) out of existence, the moral/non-moral distinction is in danger of *conceptual* collapse.

Let me repeat this point. We can dismiss feminist objections to our defining female morality (e.g. that of mother-love) out of existence by pointing out that the ancients thought of whales as fishes, whereas we prefer to think of them as mammals. Presumably this is because for them the regularities to be captured by 'things that swim in the sea' was more useful than those that can be captured by 'things that feed their young milk'. Similarly, then, we can appropriate the word 'moral' to 'responses to and only to mixed-motive games' provided this appropriation enhances our understanding of what will *now* count as moral phenomena without damaging our understanding of phenomena now no longer subsumed under that rubric. But what we've now seen is that even by our *own* lights the moral/non-moral distinction is in danger of collapse. Can this collapse be prevented?

Well, we *could* say that a disposition is moral just in case it has the logical properties of a rational intra-mental response to a mixed-motive game and *it has some such instrumental history*. But for those of us who subscribe to the Five Minute Hypothesis this would render virtually *all* our dispositions non-moral.² So perhaps we should say a disposition is moral just in case it has the requisite logical properties and it *could* have had such an instrumental history. But this won't work either, since there's virtually no disposition for which some such story *couldn't* be told. Then perhaps we should say a disposition is moral just in case it has the requisite logical properties and *in its absence the agent would fare worse than she would otherwise*.

What's just entered, note, is ...

4. Evolutionary Ethics

The evolutionary ethicist is a naturalizer of discourses. Unlike the phenomenologist, who eschews any analysis of experience that doesn't take seriously the *subjectivity* of experience, or the personalist who insists on the ineliminability of person-talk, or the anti-reductionist about whatever, the evolutionary ethicist bears no loyalty to the folk-categories by which the hoi polloi manoeuvre their way through the material and social world. The moral/non-moral distinction, says she, is a *folk*-distinction. And, she adds, a gratuitous one at that. So, says she, let's just say that:

There are certain memes - memes triggered by mixed-motive interactivity - whose capital-N-normal³ functions are such that, having invaded the meme pool, they've

² Come to think of it, for those looking for a thought-experiment to capture Kantianism, look no further! Kantianism is just the view that all truly moral dispositions are in just this sense non-moral!

³ The capital-N normal function of a feature - according to Ruth Milliken, who coined the phrase - is that in virtue of which the genetic coding for it survived in the gene pool. One criticism of this understanding is that it leaves features with far too many functions. For example, is the function of lips to keep the edges of our mouths from fraying?

replicated there because they've proven reproductively satisfying to their hosts.⁴

Or, what amounts to the Five Minute Hypothesis version of the same thing:

There are certain memes which are triggered by mixed-motive interactivity and which are such that, having *made up* the meme pool, they *will* replicate there because they *will* prove reproductively satisfying to their hosts.⁵

But this won't quite do either. For if ours were the last generation on earth, on this analysis there could be no such thing as morality. Furthermore, suppose we selected a meme because we had good grounds to *anticipate* it would maximize our take in mixed-motive games, but the world consistently conspires against us, in much the way that it consistently conspires *for* us on the God's-a-Mother-Hen Hypothesis. So to repair the definition against *both* these counterexamples, let's say instead that a moral meme is just:

a meme, triggered by mixed-motive games, which is *either* a) naturally selected for because of its reproductive fit, or else b) deliberately designed to maximize on the satisfaction of the agent's preferences.

One advantage of this analysis of the scope of morality is that it renders the Creationist/Darwinian debate moot. Why do we behave morally? On both accounts, because we have an algorithm that so directs us. Why do we have such an algorithm? On both accounts because otherwise we couldn't get along. The *historical* counterfactualist, i.e. the Darwinian, wants to add, "And if we hadn't been able to get along we wouldn't be here!" But this is gratuitous. The point is that if we don't figure out - from game theory, Scripture, or what have you - *how* to get along, we won't be here much longer!

The difference between the two views - in fact the *only* difference that could *make* a difference - is that stories about how we *came* to have the moral dispositions we have tend to cite material conditions that constrain our plasticity, and so constrain which options are worth *deliberating* about. For example, suppose that, because of our evolutionary past, partiality to our own offspring is virtually hardwired. Then whether such partiality is or isn't instrumental to our getting along now or in the future, only at the cost of considerable biological kick-back do we mess with the institution of inheritance. But, apropos our earlier concerns about the moral impact of naturalization, this does *not* say we shouldn't mess with the institution of inheritance. It says only that we should *inform* ourselves about what we're messing with before we mess with it!

Probably the most universal experience that exemplifies the relevance of evolutionary history is lust. Creationist ahistorical morality tells us *not* to lust. What it does *not* tell us is why we need to be *told* this. An evolutionist morality, by contrast, tells us *why* we lust. It tells us what natural condition the "Don't lust!"

⁴ Or, cf. Sober and Wilson, to the group of which these hosts were members.

⁵ Or, cf. Sober and Wilson, to the group of which these hosts were members.

commandment is designed to combat.

One could suppose that the *devil* makes us lust. But, once again, this merely begs the question. Why does the devil want us to *lust*? Because lust is deleterious to human flourishing? But in many if not most cases this is false. So we need to know under what conditions lust is deleterious and under what conditions it's not. But one can't answer that question by consulting God and/or the devil. One can only answer it by an *independent* assessment of the relationship between lust and human flourishing. And by an independent assessment of what *counts* as human flourishing.⁶ Evolutionary ethics can't *provide* that assessment, but it can lend invaluable assistance.

As already noted, game theory is to evolutionary ethics what grammar is to language. It tells us what *structure* an evolutionary ethical explanation must satisfy to count *as* an explanation. So, that said, back to game theoretic apparatus and, more particularly ...

III. EXTRA-LOGICALITY

Let's say that by a purely *logical* feature of a social problem we'll mean one in virtue of which we identify its game theoretic *type*, whereas by an *extra*-logical feature of it we'll mean one that's extraneous to its form but is nonetheless key to its solution. So, for example, the ordinals of our preferences having the structure they do is a *logical* feature of a Prisoners' Dilemma, but the cardinals making up its payoff matrix, the availability (or not) of an externalist solution, like a Hobbesian Sovereign, and so on, are *extra*-logical features of the situation.

Jonathan Swift made mock of his countrymen by comparing the civil wars of England to those between the Big-Enders and Little-Enders. The fact that eggs are vertically asymmetrical is clearly an extra-logical feature of the world. But what role is that fact playing in the conflict between the two Lilliputian factions? Swift doesn't tell us. Nor does he think it important to ask. But it is. For the like can be said - can it not? - of the *filioque* between Croats and Serbs.

Croats are Catholic, Serbs Orthodox. Catholics believe the Holy Spirit "proceedeth from the Father *and the Son*" - in Latin, *filioque* - whereas Orthodoxy holds that the Holy Spirit "proceedeth from the Father" alone. According to Orthodoxy, then - but not Catholicism - the Holy Spirit enjoys an ontological status equal to that of the Son. And this explains why Orthodoxy, as distinct from Catholicism, has such a high and well developed Pneumatology.

Fair enough. But how many Croat or Serb peasants-turned-soldiers could have told you this? Precious few. If asked the difference between his people and 'theirs', more likely the Croat would answer, "We're the ones who have *one* bar on our cross, whereas those Serb bastards have two!" So, one might argue, here's what was *really* going on in the former Yugoslavia:

There's always been stiff competition in the Balkans for resources, most especially land. So we need to kill off - or at least expel - x percent of the population. Arbitrary killing is a highly unstable culling strategy. Hence we need to identify a *genus* making up x percent of the population. In other words, it's not that we're hell bent on *ethnic* cleansing, or on committing *genocide*. We're bent on committing x-

⁶ Come to think of it, isn't this all and precisely what Plato was trying to tell us in the *Euthyphro*?!

perenticide! But since the Balkan population is *phenotypically* homogeneous, we need to *uniform up*, so to speak. Hence the two cruciforms.

Likewise, then, must it have been among the Lilliputians. Swift can make mock of the Big-Enders/Little-Enders distinction - and by implication the conflict between Catholics and Protestants in England - only because he's taking the rhetoric of the Catholic/Protestant conflict at its word. Similarly, the conflict between government and rebels in Colombia isn't over ideology, there being no ideological distinction that could survive scrutiny. It's over control of the drug trade. The conflict between us Jews and you Arabs in Palestine isn't over the Abrahamic Covenant. Nor is it about my rapacious mother Sarah's insecurities over our father's affection for your mother Hagar. It's about real estate and water. There's not enough of either to go around!

But, comes the objection, even supposing much of the Hebrew Bible is designed to forge and maintain a looting coalition, the two cruciforms didn't arise out of some need to bifurcate the population! Marxism wasn't an artifice of the drug cartels.

True enough. But, enter the Five Minute Hypothesis. What we want to know is not how the cruciform distinction became available. What we want to know is what work that distinction is doing for us *now*. For it's in virtue of our understanding of *that* work - not the work it once did - that we're in a position to predict whether it will or won't survive, and/or to determine whether we might replace that distinction with one that does the same work but with less destruction. So ...

IV. EXTRA-LOGICALITY AND RACISM

1. Racism, Looting, and Symmetry-Busting

Here, then - and at long last! - is just a representative sampling of the work *racism* might be doing for us:

A couple weeks after I was married, my then-wife and I were at a dinner party, the subject of interracial marriage came up, another guest made some observation, and my then-wife said, "Yes, that's what we find too, don't we, Paul?"

"Huh?" I enquired.

"You know. Us. That's what *we* find!"

"Excuse me," said I, "but what color do you *think* I am?"

"Well," she replied, "you're certainly not *white!*" Which was, well, the first I'd heard of it!

"Look," said I - remembering both my highschool social studies class, in which we were told the world was divided into four races: Caucasians, Negroids, Mongoloids, and Australian Bushmen, and remembering the (albeit controversial) work of University of Western Ontario Psychology Professor Philippe Rushton, according to whom blacks are less intelligent but they have larger penises, orientals are smarter but they have smaller penises, and whites are halfway in between, on both measures⁷ - "Look," said I, "much as I'd like to be, you *know* I'm not black. But hey, grant me this: I'm not exactly oriental either! So what's left? Are you suggesting I'm Australian Bushman?!"

⁷ This is, I concede, a bit of a caricature of Rushton's 'research'. But it was more or less what he was *taken* to have said, and gave rise, needless to say, to considerable furor.

What explains my then-wife's confusion - or, perhaps, mine - is that she's the daughter of an Anglican vicar. According to upper-class Brits, the white/non-white distinction runs pretty much where Occupied France met the Vichy Republic. But I've told this story - much to her chagrin, and so as often as I can - because it highlights just how plastic our racial categories can be. Still, the great virtue of race is that once a race is acknowledged *as* a race, membership *in* it seems to be highly *determinable*, and so *highly resistant to dissimulation*. This makes it an ideal candidate for *symmetry-busting*. The availability of a symmetry-buster, as we're about to see, is often the sine qua non of coalition-forging. And coalition-forging is sine qua non to the principal means by which human beings have always, do now, and always will advance their interests, namely looting others of their species. To explain:

Consider a three-party looting game. Call it *The Good, the Bad, and the Ugly*. If Good has 10 widgets, Bad 6 and Ugly 2, what's the rational coalition-vector for Bad? Obviously Ugly, since, assuming a fifty-fifty split of the proceeds, Bad's total wealth is now 11. Likewise should Ugly join forces with Bad, since otherwise he'd net only 3 instead of 5. The rational strategy for Good, therefore, is to *divest* herself of 4 widgets, preferably by giving all of them to Bad. And so Bad, of course, would immediately have to pass them on to Ugly. Thus looting games in which there's a *prior* asymmetry of endowments are excellent models for analyzing taxation.

Fair enough. But what's the rational strategy when there's no such prior asymmetry? Or when looting has *already* produced a symmetry of wealth? Suppose, that is, that Good, Bad, and Ugly each have 6. As Good I reason that a coalition with Bad makes as much sense as a coalition with Ugly. Unfortunately, however, for Bad a coalition with Ugly makes as much sense as a coalition with me. And likewise for Ugly with respect to Bad. Arbitrary coalitions, like arbitrary culling, are inherently unstable. So what we need is some kind of *salient* solution. Preferably one with some 'stick' to it!

Now suppose some *extra*-logical feature, like Good and Bad being of one color and Ugly being of another. Problem solved? Not at all. For the symmetry persists. How so? Because 'Ally with like!' and 'Ally with unlike!' are themselves symmetrical injunctions. So now I need to bust the symmetry between these symmetrical symmetry-busters!

But, fortunately, there *is* a way to bust some symmetrical symmetry-busters. For though it's true that from the perspective of each of the three players there are no asymmetries to be found, from the *God's eye* point of view, so to speak, there are. From the God's eye point of view - but only from the God's eye point of view - there's a second-order asymmetry between the one all-white coalition and the two mixed-race ones. That second-order asymmetry is that there are two of one kind but only one of the other. So the all-white coalition is salient. So here's a possible error-theoretic account of racism:

Racism - the view, albeit mistaken, that race has significance *beyond* its utility as a symmetry-buster, and then the further, albeit mistaken, inference that it's this transcendent significance that *grounds* the symmetry-busting it performs - is parasitic on the capacity, a capacity arguably confined to human beings, to take that God's eye point of view. If this two-part error theory is correct, it should not surprise us that, though other species can exploit arbitrary 'racial' distinctions among themselves, only humans are capable of *racism*. Nor should it surprise us that our racist attitudes have the phenomenology of transcendental truth!

But enough of Hegelian logic. The claim, in a nutshell, is that race is a for-all-our-intents-and-purposes natural kind which can be exploited for symmetry-busting for resolving coalition impasses in

looting games, and that *racism* is the psychological mechanism required to stabilize this choice of busters. Now let's see if this analysis does any work.

Imagine a situation - hell, don't imagine, just look to basic economics! - in which we'd all be better off, including those now rendered relatively *worse* off, if some of us were rendered *better* off than others. The view that we should opt for this asymmetry is called, by economists and game theoreticians alike, *paretianism*. Most political theorists, foremost among them John Rawls, are paretianists when it comes to divvying up the *material* dividends of civil society, but not when it comes to *political* dividends. Others, myself included, hold that material and political dividends are of a piece. They're all ultimately reducible to the common canonical currency of preferences. Thus the blacks of Mozambique who were trying desperately to get *into* South Africa during Apartheid - notwithstanding that if successful they'd forfeit the *political* equality they enjoyed in Mozambique - were, on my view, but not Rawls', acting rationally.

Imagine, then, that we'd all be better off - including the slaves - if x-percent of us were slaves. Or, to simplify the thought experiment, imagine the Martians had invaded and rendered all of us slaves, but they're prepared to release all but x-percent of us if, but only if, we all agree, including those who *won't* be liberated, to this partial liberation. And, suppose, as already noted, the liberation of the soon-to-be-non-slaves will accrue in at least *some* measure to the improvement of the conditions of the still-slaves. Now suppose further that x is precisely the percent of both blacks in the population and blue-eyed people in the population, that no *other* 'cut' is available - or at least none that will be countenanced by the Martians - and that neither will they allow non-slaves to *compensate* slaves. So we have full symmetry between black slavery and blue-eyed slavery.

Now then, suppose we flipped a black and blue coin and it comes up black. Would the continuing enslavement of blacks constitute racism? Clearly not. Would it make a difference if instead of only some of us being liberated the Martians have *just* invaded and so only some of us are to *be* enslaved? Hard to see how. And so, finally, now suppose that a condition of maintaining this institution of race-based slavery is that we need to self-efface our acknowledgment of the radical contingency of the enslavement of blacks rather than blue-eyed people, and replace this acknowledgment with some metaphysical nonsense about the innate inferiority of blacks. If the race-based slavery was justified, why not the *racism* required to maintain it?!

It seems to me that the anti-racist has available to her only two avenues of escape. Either she can reject the game theoretic reductionist schema tout court - in which case we're two ships passing in the night - or else she must allow that under such conditions racism *would* be justified. But, she must add, such conditions *have* never been realized in the real world, *are* nowhere in the real world realized, and *will* never in the real world be realized.

The cost of this concession, however, is considerable. For what she's conceding, in essence, is that racism is only *contingently* indefensible. The concession is isomorphic to that forced on the pro-Choice advocate. If a plague were to render all but a half dozen women in the world infertile, would anyone bother claiming that a woman has a right to control her own reproductivity?! So that right, though it's unarguably hers in the United States, Canada, Australia, New Zealand, and Western Europe in 2003, is nonetheless a contingent right. And yet the acknowledgment of the contingency of a rights-claim, though no surprise to philosophers, is a rhetorical *disaster* in the sturm and drang of real world gender and race politics!

Is there a way out of this result? I suspect not. Need we lose any sleep over it? Would that we

needn't, but I think we do. All over the world people are being held in variously dire conditions of subjugation on the grounds that any feasible alternative is pareto-inferior. Child labor - including child *sexual* labor - is a case in point. And, it seems, we *buy* the argument, just as we buy the products of that labor. Something is amiss. But what?

2. Racism as a Solution to an Opacity Problem

We've just seen that the symmetry-busting analysis of racism has a highly unsavory upshot. Let's see if we can fare more palatably with the *readability* interpretation.

On the coalition/looting view, recall, racism is the result of just one of any *number* of extra-logical features of the world that can be strategically exploited in the pursuit of ends which are anything but mysterious, namely coalition-forging for the purposes of looting. Likewise, then, is race only one of a *set* of reactions to impediments to our ability, in this case, not to *loot* each other but simply to get to *know* each other. That is, many of our cooperative dispositions are *conditional* upon the cooperative disposition of our co-players. But this conditionality requires that players be - ideally, of course, *transparent* to each other, but even in the real world - minimally *translucent*. In other words still, we need to be able to read each other, but radical racial dissimilarities can be an *impediment* to readability.

In Gauthier's model, *Morals by Agreement*, there are only two dispositions, Unconditional Defection (**UD**) and Conditional Cooperation (**CC**).⁸ Peter Danielson's *Artificial Morality* adds Unconditional Cooperation (**UC**) and Reciprocal Cooperation (**RC**), the difference between **RC** and **CC** being that, whereas **CC** generously cooperates with **UC**, **RC** ruthlessly exploits her.⁹ What's important for *our* purposes, however, are the scrutiny costs borne by **CC** and **RC**, the *autophany* costs borne by all but **UD**, the opportunity for *dissimulative* autophany this need to read each other presents, and *differential* costs that might arise out of players' *extra-logical* features, such as, well, race.¹⁰ How do you tell when a black man's blushing? Answer: you and I can't. At least not without getting to know him for a tad longer than it takes to read a litter mate, so to speak. But remember: even a few seconds can make the difference between seizing and foregoing a cooperative opportunity!

But, as I say, race is only one such feature. I speak only two languages, but I can tell from the timbre and lilt of virtually any Indo-European language whether the speaker's angry, sad, happy, joking, or what have you. Not so with Mandarin or Cantonese, even after a week on the Chinese mainland. So I shrink away. Why do banana republic dictators wear one-way shades? Because they can see your eyes but you can't see theirs. Why do torturers strip their victims? Do clothes offer that much protection against the heft of a truncheon? No, it's that the body gives itself away. Why am I uncomfortable leaving a message on your answering machine? Because you've had hours to perfect what you reveal to me, whereas I'm left to stutter my way through what revelation is being required of me. In short, human beings abhor being on

⁸ David Gauthier, *Morals by Agreement*, Oxford U.P., 1986

⁹ Peter Danielson, *Artificial Morality*, Routledge, 1992

¹⁰ I coined the word 'autophany' in my [omitted for vetting purposes].

the short end of any asymmetry of visibility, because information is power. Readability across radical racial dissimilarities is doable but expensive. Thus it shouldn't surprise us that we tend to prefer interactions with those we can more readily read. And, of course - and here's the error theory attendant on the readability interpretation - *racism* is just what one would expect of an unreflective reflection on *why* the 'other' is so difficult to read. They're not readable because they're, well, psychopaths, or not fully human, or whatever.

This species of 'justified' racism, though no less implausible than the coalition/looting interpretation, at least has the virtue of being less offensive. No less implausible but less offensive still is ...

3. Racism as a Solution to a Vulnerability Problem

The exposition of this third hypothesis requires closer attention to the replicator dynamics among the four dispositions just introduced:

What Danielson's *Artificial Morality* purports to show is that, provided scrutiny costs are not unreasonably high, and provided **RC** is required to pay no higher scrutiny costs than **CC**, **RC** is categorically equal or superior to any of its three competitors. Populations reach equilibrium when an invader can be indifferent with respect to which disposition to adopt provided that disposition is extant in the population about to be invaded. And, as often as not, that equilibrium includes **UC**.

Such equilibria, however, are often anything but stable. Individuals/sub-populations can be *dislodged*. And almost invariably the first to go is **UC**. This is because **UC** enjoys *no* protection against exploitation. So, it might be thought, *anything* that might trigger defection is better than no such trigger at all. So - and for the leg-work on this I'm entirely indebted to Malcolm Murray¹¹ - divide the population into not four but *eight* types, prefixing each of Danielson's canonical four with either a **W** or a **B**. Now take each of these eight and add either an **R** for racist or an **N** for non-. (As we'll see momentarily, **N** acts only as a dummy place-holding operator.) So we have a total of *sixteen* types. Players behave as the last two letters in their algorithm dictate, save that white racists don't cooperate with blacks, be they racist blacks or not, and black racists don't cooperative with whites, be they racist whites or not.

Now then, using the standard reverse-ordinal scoring system¹², and even putting scrutiny costs aside, it turns out that **RWUC** and **RBUC** fare better than the **NWUC** and **NBUC**. *Quod erat demonstrandum*: for at least *some* populations, *some* payoff structures, and *some* dispositions, racial intolerance pays higher dividends than racial indifference!

4. Racism, Anti-Racism and Anti-Anti-Racism

Until, that is, we consider the effects of *anti*-racist dispositions. The anti-racist white **RC**, for example - let's call her **ANWRC** - *would* cooperate with, say, the racist white **CC**, were it not for the latter's racism. The idea here is isomorphic to the game-theoretic model of the New York subway system, consisting of

¹¹ Murray's findings are as yet unpublished.

¹² That is, 4 for unilateral defection, 3 for mutual cooperation, 2 for mutual defection, and 1 for unilateral cooperation.

hawks, doves, and resisters. Hawks are muggers, doves mugges, and resisters Guardian Angels. We're *inclined* to think, at least intuitively, that the latter disposition must be an altruistic one. And we might be inclined to think likewise of anti-racists. But we'd be wrong. On both counts. Just as we'd be wrong to think of **RC** as predatory. The *true* predators among Danielson's canonical four are the **UDs**. **RCs** kill them off by killing off their food supply, namely the **UCs**. Likewise, though it *appears* that resisters are altruistic, the fact is they fare quite well, so long as the equilibrium's just right. And, as it turns out, the same can be true of anti-racists. In short, the substantive results of an interactive ecology cannot be anticipated by mere hunches. As Danielson is wont to say, "Forget your intuitions. Do the math!"¹³

But, of course, if there can be racists and anti-racists, likewise can there be - and there are! - *anti-anti-racists*. In fact one needn't even be a *racist* to be an anti-anti-racist. For remember: when doing virtual ethics, what matters, and *all* that matters, are the relative payoffs of our dispositions, not their rationales. Nor even whether these rationales are logically consistent. That, recall, was the lesson we learned from the God's-a-Mother-Hen Hypothesis.

Nor is this toleration of inconsistency a purely *theoretical* observation. For example, I'm having a theodical tiff right now, with a God whose existence I deny. Many of my students find this odd. Until, that is, I point out to them that if haecceitism is true then there might well be an infinite number of each of us, which is absurd, but if it's false, then if you drive a spike into an absolutely symmetrical ball, it comes out the same side you drove it into, which is even more absurd.¹⁴ So, we all believe that haecceitism is both true and false. In fact one *has* to believe that haecceitism is both true and false to have a web of beliefs that will prove in any wise serviceable! What game theoretic replicator dynamics show us is *why* logical inconsistencies are no cause to lose any sleep! Memes replicate by serviceability, not consistency!

In fact, notwithstanding that I'm convinced I'm not a racist, I think I *am* an anti-anti-racist. Am I the only one who finds anti-racists morally smug?! Nor would it surprise me if I discovered I *was* an anti-anti-racist, notwithstanding my not being a racist. The dynamic might be akin to the following:

I'm not a rapist. But neither am I an *anti-rapist*. I might, for all I know, be an *anti-anti-rapist*. Why? Because - and feminists have been making this connection for decades! - much as I'd never be a rapist myself, and much as I don't want my *own* partner to be raped, I want there to be just enough rapists out there that my partner must cleave to me for protection. For this dependence gives me power in our relationship that I otherwise wouldn't have.

As I say, feminists understand this. In fact, though they may recoil at the conceptual framework in which I've explicated the phenomenon, they're the *first* to point out that the replicator dynamics for patriarchy depends on neither ideological consistency nor conscious solidarity. It's all about payoffs, payoffs, payoffs! So, likewise might it be with racism.

Of course one can always concoct a model, and cook one's inputs, to prove some theoretical point. I've used a similar technique to show that, the insistence of its detractors to the contrary notwithstanding, game theory can - indeed, I argue, *only* game theory can - solve the problem of justifying

¹³ in private conversation

¹⁴ Haecceitism is the belief that there exists the property of being 'this'.

concern for future generations.¹⁵ That is, I did so by showing that “There exists a set of values such that ...” But as with the coalition/looting interpretation of racism, what’s needed here is a proof with values reflective of the world one purports to explain the emergence of racism *in*. But *that*, alas, is far beyond the scope of the present undertaking.

V. EXTRA-LOGICALITY, LOGICALITY, AND RACISM

What I *do* want to leave off with, however, are some worries I have about the distinction between the logical and extra-logical features of our models. Is there important work this distinction is doing for us? If so, what happens to that work if the distinction can’t be maintained?

Recall that I said at the outset that by a *logical* feature of an interaction I’d mean what gives it its *game-identity*, whereas by an *extra-logical* feature I’d mean something *happenstantial* about the world that can be exploited to *resolve* some coordination problem. So, for example, in a problem like walking a fridge, that most of us are *right-leg* dominant probably accounts for why, even *without* prior agreement, we all seem to heave first to the left. But right-leg dominance isn’t a *logical* feature of the world. Human beings aren’t *essentially* right-leg dominant. So, in addition to the error theories already advanced, *another* way to understand *racism* is as the taking, albeit mistakenly, of something *happenstantial* - like the emergence of *black* slavery rather than blue eyed - as reflecting something *deeper* about the nature of blacks.

What’s particularly appealing about this kind of approach to racism is that it preserves the intuitions of those of us with ‘ejacashuns’ that there’s something *wrong* with racism, while preserving the dignity of the hoi polloi by noting that racism is an *honest* mistake - or at the very least an *understandable* one. What’s *unappealing* about it, however, is that it reminds us that even those of us *with* ejacashuns aren’t sure what is and what isn’t a contingent feature of human interactivity, and so what is or isn’t grist for this kind of error theory. For example:

The earth revolves around the sun in such a manner that, for all intents and purposes, it’s *wobbling* north-south, and this *wobbling* gives rise to seasons. Doesn’t the fact that ours is a *seasonal* existence give rise to the need to store our food supply? And doesn’t the need to store our food supply give rise in turn to non-perishable tokens of exchange, which, according to Locke and Rousseau and Marx, gives rise in turn to government, and thence to war? So, it would seem, our ‘seasonality’ is the *sine qua non* of the kinds of games we play. Is that seasonality an *extra-logical* feature of the human condition, or a *logical* feature of it?

And to whom are we referring when we talk thus about ‘humans’? Are *equatorial* people as seasonal as Capricornians and Cancerians? Does that mean they’ve evolved to play different games? And if so, could it be argued that, if slavery there must be, then black slavery is *not* a mere accident of history. In fact, if the story just told is to be believed, black slavery was *virtually inevitable*.

And *this* is why, for all its wonkiness, the Five Minute Hypothesis is not to be so readily dismissed. Absent the Five Minute Hypothesis there’s no way, at least in principle, to rule out the kind of just-so story

¹⁵ Omitted for vetting purposes.

I've just been telling. The kind of story, in fact, one might expect coming out of German universities in the late 1930's and early 1940's. The Five Minute Hypothesis blocks these stories. It asks only whether black slavery could be *sustained* if it existed, or could be *instituted* if it did not, under the current conditions. And the answer to that, it seems, is clearly no.

VI. SUMMARY

So, let's sum up. My suspicion is that something as widespread and persistent as racism cannot conscionably be explained away as a completely unmotivated mistake. Accordingly, I've offered three hypotheses about its genesis: first, that it's a coalition strategy for a looting game, second that it's the product of differential readability, and third that it's a defense against vulnerability. I'm not particularly sanguine myself about any of these hypotheses, but I *do* think they should be taken seriously, because if anything *like* them is what's going on, we can *use* this understanding of racism to combat it.