The Matter of Self-Sabotage and White Racism: A Stratification Economics Perspective

William Darity Jr. and David McMillon

September, 2025

Introduction

The argument that everyone, or, almost everyone, loses from racism has become a popular perspective in political discourse, leading to the appealing "shared prosperity" perspective actively taken on in the public and private sectors (Buckman et al., 2021; Peterson, 2020; Treuhaft, 2014). One of the most prominent presentations of this line of thinking is Heather McGhee's (2021) *The Sum of Us*, where McGhee contends racism hurts everyone materially, both the victims and perpetrators. This perspective is bolstered by a litany of work showing how racial discrimination has induced various harmful distortions on the economy (Hsieh et al., 2019, Cook, 2014) which, in shrinking the "pie," also impact White Americans. This argument necessarily ascribes a certain irrationality to the practice of racism. If everyone—or virtually everyone besides a tiny White elite—is made worse off by racism there could be a quasi-Pareto improvement following its elimination. This leads to the question: if White racism is self-sabotage, why does it persist?

There are at least three major variations on the claim that White Americans' prejudices lead them to persistent self-sabotage. The first is the claim that racism persists due to psychological or cognitive limitations. For example, Whites may hold prejudicial beliefs out of misinformation or flawed inferences–which, if corrected, would allegedly reduce behaviors that are ultimately self-destructive. An example of this is inaccurate statistical discrimination, in which a firm holds inaccurate beliefs due to misinformation about Blacks in general, leading to consistent underestimation of Black productivity, ultimately harming the [White-owned] firm (Bohren et al, 2025). In a more affective sense, racism has been argued to be a

psychopathology, a mental illness that leads its practitioners to make self-destructive decisions. This appeared as early as 1927, in an intentional provocation directed against the proponents of scientific racism who declared "the Negro" congenitally deficient, E. Franklin Frazier deemed anti-black prejudices held by whites to be a form of psychopathology—a mental illness. In his article, "The Pathology of Race Prejudice," Frazier commented that his diagnosis might hold promise for positive change. If race prejudice is a disease, it might be cured; the proper prescription simply has to be found.

Second is the notion that historical social meaning-making along racial lines have generated a persistent racial stigma (Loury, 2002)–through which Whites unknowingly or unwittingly view Blackness itself without the presumption of a common humanity. Removing this more deeply-rooted social stigma would move society towards a collectivist existence that would ostensibly be of benefit to Whites in the long term. For example, so the argument goes, Whites may have benefitted from greater attention to systemic vulnerabilities to climate change and lead exposure if the victims of Hurricane Katrina or the Flint water crises were viewed subconsciously as equally human.

Third is the claim that whites—or at least the white working class—acts against its own self-interest because it has been manipulated into holding racist beliefs that preclude uniting with blacks for mutual gain. At minimum, so the argument goes, the white working class in the United States frequently votes against its own self-interest because it is influenced into vitriol for the presumed beneficial effects for Black Americans despite potential benefits for themselves. Here White racism is induced by the hegemonic practices of a White elite pursuing a divide-and-rule strategy vis-à-vis the working class.

This paper interrogates the claim that White racism is a form of "self-sabotage," clarifying when, for whom, and over what horizons such a characterization may be warranted. We use "White racism" to reference beliefs, practices, and institutions that create and reproduce systematic advantages for people socially classified as White relative to non-White groups. This definition follows public health and sociological literatures that emphasize structural arrangements—not merely individual animus—as causal mechanisms (Braveman et al., 2022; Omi & Winant, 2014).

We challenge what we call incentive-free explanations—the aforementioned attribution of persistent racist beliefs and practices to psychological, cognitive, or phenomenological distortions, and social manipulations that purportedly induce many white Americans to act against their own material interests. We argue that these explanations generally imply remedies for unintentional errors: interventions that seek to correct mistaken beliefs, recalibrate inferences, or cultivate collective empathy. Examples include curricular and public-history reforms, social contact and perspective-taking programs, social-norm messaging, and even clinical deradicalization. These policies predict limited strategic behavior on the part of advantaged actors: once better informed or more empathetic, Whites should choose less discriminatory policies. The stratification-economics lens cautions, however, that where status or material rents are meaningful, information and empathy often face materially motivated headwinds and generate fragile or domain-specific behavior change; absent incentive realignment, actors can substitute into new closure margins even as explicit attitudes soften.

We therefore also consider what we call incentive-based accounts—which treat racism as a set of beliefs, behaviors, and institutions sustained by rational (or boundedly rational) incentives: the extraction of material rents, the protection of relative status, and the maintenance of political advantage (Darity 2005. Darity 2022; Chelwa, Hamilton, & Stewart, 2022). In this

framework, what looks like self-harm at the level of aggregate efficiency may reflect "rational" attention to positional payoffs—what Du Bois called a "public and psychological wage" of whiteness that compensates some whites for foregone income in exchange for relative racial status and power (Du Bois, 1998). Under this perspective, if racism is material self-sabotage for Whites, it is because they prefer it—whether due to present bias, devaluation of material rents, or the willing protection of relative status. Rather than being manipulated per se, working class Whites become rational preference maximizers. Racist beliefs and practices can be sustained even with a mentally healthy, highly educated, self-aware, and collectively empathetic population. Incentive-based accounts imply that racism will diminish only when those payoffs are neutralized or reversed.

Using insights from stratification economics, we evaluate the empirical record and review some conceptual groundwork for formal models of utility over relative status that we develop in the remainder of the paper.

II. Incentive-free arguments: Racism as an Erroneous

Mental-illness accounts.

A long lineage—running from E. Franklin Frazier's (1927) early essay to later debates in medicine and sociology—frames racism as pathology in the racist. Contemporary commentary often invokes neuroscience to portray hatred as a brain-based disorder (Hayasaki, 2018). While racism never has been included in the *Diagnostic and Statistical Manual of Mental Disorders*, a long running current of belief in the idea of racism as a disease has been sustained among psychologists. Indeed, throughout the three decades between the 1920s and 1950s, psychologists investigating racism generally "considered prejudice to be a psychopathology" (Hayasaki, 2018),

and this belief remains embedded in much contemporary research. For example, Alvin Poussaint (2002) argued "extreme racism" is a delusional condition leading to violent and other types of dangerous actions requiring medical intervention.

While it is crucial to recognize that racism demonstrably harms the health and mental health of its targets (Braveman et al., 2022; Lewsley & Slater, 2023), the claim that racists are, as such, mentally ill has been repeatedly critiqued for medicalizing what are known social, political, and economic projects (Poussaint, 2002; Thomas, 2014; Thomas & Byrd, 2016). Treating racism as an illness directs remedies toward clinical interventions for perpetrators and, potentially, away from incentives and institutions that generate durable advantages for Whites through racism. If White racism would survive a mentally healthy populace, it could be because Whites benefit from it materially—casting doubt on the "self-sabotage" argument—or otherwise because there are other incentives, or social, cognitive, psychological, or phenomenological errors that would allow it to survive.

B. Misinformation and Flawed Inferences.

A second incentive-free line contends that racism stems misinformation and flawed inferences that can be harmful for racists. Therefore, a more informed, well-educated, and introspective public would abandon discriminatory beliefs and behaviors, in part to the benefit of former perpetrators of racism.

Yet social psychology cautions that people actively process information in identity-protective ways: motivated reasoning leads individuals to defend group-congenial conclusions and discount contrary evidence (Kunda, 1990; Kahan et al., 2017). In the presence of material or status payoffs to a racial hierarchy, better information may be insufficient; actors can

know that stereotypes are false and nevertheless propagate them because doing so helps preserve advantages. Thus, even if improved information reduces prejudice at the margins, it cannot by itself dislodge incentive-compatible structures. If White racism does lead to self-sabotage through misinformation and flawed inferences, it only does so to the extent that we ignore incentives.

C. Racial Stigma.

Racial stigma, in Loury's account, is a socially reproduced schema that marks blackness with negative attributions and lowers the intrinsic worth of Black persons in society—shaping beliefs about desert, competence, and trustworthiness and thereby channeling opportunity (Loury, 2002). Loury distinguishes this as something deeper than "simply" racism: it is the lack of the presumption of a common humanity for black persons.

This affects how one regards black suffering: 1) whether it is considered an affront to society, or an acceptable part of the natural order of the world; and 2) the attribution of responsibility–whether we attribute responsibility to the victim or to ourselves for allowing a society to exist in which this happens. This kind of deeply entrenched social *meaning*–not merely social attitudes–would require a rather grandiose project in collective empathy. It would require a socialized re-wiring through which Blacks become "us" rather than "them."

Read this way, persistent discrimination can be sustained by widely shared but deeply psychologically entrenched, institutionally reinforced stigmas that make discriminatory choices appear reasonable to decision makers at low private cost. Yet, stigma is itself endogenized by rules, markets, and policies that allocate rents and structure contact; it is reproduced not only through non-strategic socialization processes, but partly because it is useful to maintaining

boundaries. Thus, while fighting the schema of racial stigma is normatively vital, focusing on stigma without altering payoffs risks over-promising what collective empathy or catharsis can accomplish. Racial formation theory underscores the point: the meanings attached to race—and the stigmas they authorize—can be intentionally made and remade through state policy to preserve political power (Smedley et al., 2005). Moreover, "colorblind" approaches that suppress race-conscious remedies can entrench stigma's material effects by foreclosing tools (e.g., affirmative action, targeted investment) that are purported to counteract its cumulative harms.

III. Incentive-based arguments: when and how racism "pays"

A. Racism as "Economically Inefficient."

A major part of the argument behind the self-sabotage of White racism is that it is economically inefficient—that is, it suppresses innovation, productivity, and aggregate output, and causes arbitrage opportunities and market failures. There is substantial evidence: anti-Black violence reduced patenting by Black inventors (Cook, 2014); discriminatory barriers misallocated talent across occupations in ways that lowered growth (Hsieh, Hurst, Jones, & Klenow, 2019); segregation and place-based disadvantage depress mobility and human-capital formation (Chetty, Hendren, Jones, & Porter, 2020; Johnson, 2019; Durlauf, 2004). "Blockbusting," the practices of scaring White residents into selling their homes at low prices using Black actors, and selling the homes to Black families desperate for better opportunities, is the result of the fact that racism can induce market failures.

The crucial point for our purposes is that inefficiency at the macro level can coexist with rents for those positioned to capture them. For example, real-estate intermediaries who were

predominately white extracted gains from blockbusting, while black households lost housing wealth (Hartley & Rose, 2023). Similarly, carceral expansion has imposed large social costs while conferring concentrated benefits—budgetary, political, and labor-market—to actors embedded in the carceral economy (McKay & Darity, 2024; Eason, 2017). Federal data indicate that the correctional workforce is disproportionately white, a distribution consistent with stratification-congruent employment rents. Therefore, even when racism does introduce economic inefficiencies, that doesn't guarantee that Whites will be worse off materially. Ostensibly the cost of the inefficiency could be passed on entirely to non-Whites. To show that White racism is self-sabotage even materially, one must show that these inefficiencies induce costs that are also shared among Whites.

One common argument is that the costs of racism are shared not among all Whites, but particularly among Whites whose economic status is below the economic elite, described in the next subsection.

B. Class manipulation versus shared gains.

A widely cited class-based view holds that White elites stoke racial division to prevent cross-racial working-class coalitions, an argument often linked—historically—to the post-Bacon's Rebellion codification of race in colonial Virginia (Morgan, 1975). This argument begins with the presumption that White elites *do* have material incentive to maintain the oppression of non-Whites, but that working-class Whites do not–and must be compensated with a psychological incentive instead. Du Bois's account of the psychological wage formalized how elites could trade status for solidarity (Du Bois, 1935). Contemporary research on status politics helps explain why appeals to perceived status threat can mobilize advantaged-group support,

even when policy outcomes are economically costly in the aggregate (Mutz, 2018; Koenig & Mendelberg, 2025; Gest, 2016). Stratification economics reframes the puzzle: if racism preserves relative advantage, working-class Whites may not be "duped" so much as presented with a package in which status rents offset, and sometimes outweigh, foregone material gains.

A canonical example of practices that may be materially harmful to Whites but compensate relative status, is the closure of public swimming pools rather than their integration (Palmer v. Thompson, 1971). On its face this looks like collective self-sabotage in that materially, poor Whites gave up amenities to preserve segregation that they would have had access to under integration. Under positional preferences, however, destroying a shared good can be privately rational if integration would erode the status premium of exclusivity. Private provisioning then emerges to restore the benefit for those able to pay, while the costs of lost public infrastructure fall disproportionately on non-Whites and lower-income Whites–who are indirectly compensated through relative racial status preservation. The fact that segregated public goods are sometimes abandoned reveals not that Whites as a class do not value the good, but possibly that a large enough subset values the good primarily as a marker of group distinction.

It is also instructive to return to the example of mass incarceration in the case of class-based White racism. Even if mass incarceration depresses aggregate productivity and imposes costs that also harms Whites materially, it may endure because it delivers concentrated material and status rents to pivotal White constituencies. Privatization converts punishment into a revenue stream for private prisons and a web of contractors. The broad "collateral consequences" that harm communities—including some White taxpayers and low-income Whites—may be outweighed politically by "psychic rents" in the form of group-status preservation and localized gains to White workers, firms, and officeholders (McKay & Darity, 2024; Western & Pettit, 2010). Thus, what

appears as collective self-denial again resolves into a distributive trade-off: elites and strategically positioned White communities capture rents through carceral expansion (public or private), while Black communities absorb the most direct harms and poorer Whites tolerate absolute losses for the sake of maintaining a relative racial status premium (McKay & Darity, 2024; Eason, 2017).

Whether racism is harmful to Whites requires showing not only that economic inefficiency costs are passed on to Whites, but also which White subgroups, and more importantly, depends on how "harm" is defined. What is harmful materially may not be harmful in a Welfare sense when utilities depend on relative racial status preferences. Psychic rents can be difficult to measure, and to our knowledge, the measurement of preferences for relative racial status advantage remains an untapped area of economic research.

However, it is still possible that racism is harmful to Whites even in utility terms in the long run (McMillon, 2025). Suppose, for example, that uncertainty and time-inconsistent preferences, as discussed in behavioral economics, lead White Americans to reject equitable policies that would improve their material well-being so much that it would compensate their preferences for relative advantage in the long run. In such a case, racism is still boundedly rational, and "incentive-free" policy solutions would not eliminate racism. Appropriate policy solutions would need to consider advances in behavioral economics for reducing perceived uncertainty and "self-control" issues– such as regularly paying projected dividends from equity-focused interventions in the shorter run, insurance, and money-back guarantees.

IV. Why "self-sabotage" persists—and why the label misleads

Calling racism self-sabotage is descriptively tempting when we observe Whites supporting policies that may harm their material conditions. But the label obscures heterogeneity in benefits and time

horizons. First, the group of "Whites" is internally stratified; some subsets (e.g., elites, those in racially advantaged labor queues) harvest clear rents while others receive smaller, largely psychological wage components (Du Bois, 1998/1935). Second, positional payoffs are relative by construction; actors may prefer a lower absolute income with higher relative rank to a higher absolute income with lower rank. Third, time-scale matters: practices that depress innovation and growth may be sustained if the near-term rents to advantaged groups exceed discounted long-run losses. Hence the mistake is not merely empirical but diagnostic: if we infer that racism persists because it is a psychological, social, or cognitive error, we may design remedies that ignore incentives to preserve relative status advantages (Darity et al., 2017; Chelwa et al., 2022).

For the remainder of the paper we turn to the concern of modeling relative racial status preferences in utility functions that may represent distinct social theories regarding racial hierarchy.

V. Specifying utility functions (Formal Modeling)

[Forthcoming]

VI. Conclusion

Stratification economics helps reconcile two truths: racism imposes large social costs, often including costs on many Whites; yet it also generates rents—material and psychological—that make persistence rational for others. Incentive-free explanations; incentive-free accounts that treat racism as an error can miss how institutionalized status competition sustains hierarchy even under aggregate inefficiency. The appropriate policy corollary for reducing White racism is not merely to correct psychological, cognitive, sociological, and phenomenological errors, but—to the extent possible—to redesign institutions so that racial exclusion no longer confers advantage, material or

otherwise. That is, to design systems such that, even in a future-discounted Welfare sense, White racism is truly self-sabotage.

References

Buckman, Shelby R., Laura Y. Choi, Mary C. Daly, and Lily M. Seitelman. (2021). The Economic Gains from Equity. Brookings Papers on Economic Activity, Fall. 71-111.

Peterson, D., Mann, C. (2020). Closing the Racial Inequality Gaps. Citi Global Perspectives and Solutions.

Treuhaft, S., Scoggins, J., & Tran, J. (2014). The equity solution: Racial inclusion is key to growing a strong new economy. Research Brief. Oakland, Calif.

Braveman, P. A., Arkin, E., Proctor, D., Kauh, T., & Holm, N. (2022). Systemic and structural racism: Definitions, examples, health damages, and approaches to dismantling. Health Affairs, 41(2). https://doi.org/10.1377/hlthaff.2021.01394

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic opportunity in the United States: An intergenerational perspective. Quarterly Journal of Economics, 135(2), 711–783. https://doi.org/10.1093/qje/qjaa051

Cook, L. D. (2014). Violence and economic activity: Evidence from African American patents, 1870–1940. Journal of Economic Growth, 19(2), 221–257. https://doi.org/10.1007/s10887-014-9102-4

Cramer, K. J. (2016). The politics of resentment: Rural consciousness in Wisconsin and the rise of Scott Walker. University of Chicago Press.

Chelwa, G., Hamilton, D., & Stewart, J. B. (2022). Stratification economics: Core constructs and policy implications. Journal of Economic Literature, 60(2), 377–399. https://doi.org/10.1257/jel.20211690

Darity Jr., W. A., Hamilton, D., & Stewart, J. B. (2015). A tour de force in understanding intergroup inequality: An introduction to stratification economics. The Review of Black Political Economy, 42(1–2), 1–6. https://doi.org/10.1007/s12114-014-9201-2

Darity Jr., W. A., Hamilton, D., Mason, P. L., Price, G. N., Dávila, A., Mora, M. T., & Stockly, S. K. (2017). Stratification economics: A general theory of intergroup inequality. In F. R. Harris, A. Flynn, S. R. Holmberg, D. T. Warren, & F. J. Wong (Eds.), The hidden rules of race: Barriers to an inclusive economy (pp. 35–51). Cambridge University Press.

Durlauf, S. N. (2004). Neighborhood effects. In J. V. Henderson & J.-F. Thisse (Eds.), Handbook of Regional and Urban Economics (Vol. 4, pp. 2179–2242). Elsevier.

Du Bois, W. E. B. (1998). Black reconstruction in America, 1860–1880. Free Press. (Original work published 1935)

Eason, J. M. (2017). Big house on the prairie: Rise of the rural ghetto and prison proliferation. University of Chicago Press.

Frazier, E. F. (1927). The pathology of race prejudice. The Forum, 77(6), 856–862.

Gest, J. (2016). The new minority: White working class politics in an age of immigration and inequality. Oxford University Press.

Hartley, D., & Rose, J. (2023). Blockbusting and the challenges faced by Black families in building wealth through housing in the postwar United States (Chicago Fed Working Paper 2023-02). Federal Reserve Bank of Chicago.

Hayasaki, E. (2018, November 27). The pathology of prejudice: What neuroscience tells us about the persistence of hatred. The New Republic.

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and U.S. economic growth. Econometrica, 87(5), 1439–1474. https://doi.org/10.3982/ECTA11427

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2025). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*, *107*(3), 605-620.

Johnson, R. C. (2019). Children of the dream: Why school integration works. Basic Books.
Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. Science, 356(6339), 532–535. https://doi.org/10.1126/science.aah6524

Koenig, T., & Mendelberg, T. (2025). The symbolic politics of status in the MAGA movement. Perspectives on Politics. https://doi.org/10.1017/S1537592725102090
Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lewsley, J., & Slater, R. (2023, May 31). What are the effects of racism on health and mental health? Medical News Today.

McGee, D. (2025). Exploitation through racialization. Quarterly Journal of Economics, 140(2), 1581–1631. https://doi.org/10.1093/qje/qjaf007

McGhee, H. (2021). The sum of us: What racism costs everyone and how we can prosper together. One World.

McKay, T., & Darity, W. A., Jr. (2024). Who benefits from mass incarceration? A stratification economics approach to the "collateral consequences" of punishment. Annual Review of Law and Social Science, 20, 11.1–11.22. https://doi.org/10.1146/annurev-lawsocsci-041922-033114

McMillon, D. (2025). Does Racial Hierarchy Really Harm Everyone? Relative Status Envy and the Economics of Reparative Reforms. SSRN Working Paper.

Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental justice. *Annual review of environment and resources*, *34*(1), 405-430.

Morgan, E. S. (1975). American slavery, American freedom: The ordeal of colonial Virginia. W. W. Norton.

Mutz, D. C. (2018). Status threat, not economic hardship, explains the 2016 presidential vote. Proceedings of the National Academy of Sciences, 115(19), E4330–E4339.

Omi, M., & Winant, H. (2014). Racial formation in the United States (3rd ed.). Routledge.

Palmer v. Thompson, 403 U.S. 217 (1971).

Poussaint, A. F. (2002). Is extreme racism a mental illness? Western Journal of Medicine, 176(1).

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American psychologist*, *60*(1), 16.

Thomas, J. M. (2014). Medicalizing racism. Contexts, 13(4), 24–29. https://doi.org/10.1177/1536504214558213

Thomas, J. M., & Byrd, W. C. (2016). The “sick” racist