Rawls and Racial Justice

Elvira Basevich

University of California, Davis

(Draft for *Oxford Handbook on the Philosophy of John Rawls* edited by Christie Hartley, Blain Neufeld, and Lori Watson. Comments welcome.)

Abstract:

This chapter explores the conceptual relation of facts about racial injustice to two key aspects of Rawls’s ideal theory. First, it explains why Rawls excludes race from his representation of a well-ordered society and why he believes this exclusion does not mean that justice as fairness cannot support racial justice. Second, it considers three recent accounts of the justificatory role of facts about racial injustice in justice as fairness, focusing on the methods of the Original Position and Reflective Equilibrium. It concludes with a reinterpretation of the method of Reflective Equilibrium that shows why it provides a promising pathway for showcasing the explicit role of facts about racial realities in ideal theory. Namely, their consideration can add justification to the idea of justice as fairness.

Key words: race, gender, racial justice, ideal/nonideal theory, justification

Introduction:

John Rawls never substantively theorized racial justice. As the late Charles Mills documents, Rawls wrote and said little about race. “Thirty years after [*A*] *Theory* [*of Justice*], he still had not moved on to race—surely one of the most pressing, if not *the* most pressing, issues of justice in the American polity. What was keeping him” (2017, 155)? Given Rawls’s relative silence on race in his published works, Mills argues that Rawls’s theory of justice cannot yield a satisfactory theory of racial justice. Rawls’s “body of work” simply “does not extend to the subject of racial justice” (2017, 162). This is a serious indictment as Mills argues that racial injustice should be among political philosophers’ chief concerns and that the idea of justice as fairness is inhospitable to a racial justice agenda.

Among his many charges against Rawls’s theory of justice, Mills objects that Rawls’s theory is an “ideal ideal,” rather than a “rectificatory ideal.” An “ideal ideal” signifies an “‘ideally just’ […] society without any previous history of injustice,” whereas a “rectificatory ideal” “means a society with an unjust history that has now been completely corrected for. Rawls really means the former, not the latter” (Mills 2017, 140). In Mills’s view, the defect of “ideal ideal” theory is that it aims to “prevent” future injustice from occurring. But “prevention generally differs from cure” and Mills is interested in a “cure” (2017, 140). Mills concludes that Rawls’s theory is ill-suited to redress racial injustice because it is not designed for this purpose. While there are many ways to interpret Mills’s objection to ideal theory, in this chapter, I survey how to respond to his claim that Rawls’s ideal theory lacks a suitable *epistemic and evaluative orientation* towards understanding and condemning racial realities. On this way of reading Mills’s objection, ideal theory is an “ideal ideal” in that it does not consider, nor is it developed in reference to, racial realities. To answer this version of the objection requires showcasing the role, if any, that facts about racial injustice must play in ideal-theory construction.

Rawls and Rawls scholars have provided a variety of responses to the criticism that justice as fairness is insensitive to the circumstances of racial injustice. This chapter surveys how Rawls and Rawls scholars treat facts about racial injustice in ideal theory. On the one hand, this chapter defends ideal theory’s empirical insensitivity to racial realities with respect to Rawls’s depiction of a well-ordered society. On the other hand, it considers the positive role that facts about racial injustice nevertheless might play in the Original Position and in Reflective Equilibrium. It concludes that an account of political knowledge about injustice, specified with the aid of Africana philosophy and critical philosophy of race, should be integral in the philosophical construction of the public standpoint of justice. Specifically, it argues that such broad-based knowledge should support the establishment of wide reflective equilibrium. Rawls’s ideal theory is not a vacuous “ideal ideal,” as Mills had worried. On the contrary, the process of achieving wide reflective equilibrium involves an epistemic reflectivity about the circumstances of racial injustice, which can add justification to the idea of justice as fairness and capture its attunement to racial justice concerns.

Section I reviews Rawls’s discussion of ideal theory and why facts about racial injustice are excluded from the idea of a well-ordered society. Section II considers the proper place of facts about racial injustice in ideal theory. It assesses three options that philosophers pursue to incorporate knowledge about racial realities to justify the principles of justice as fairness. It considers (a) the role of Rawls’s personal intentions and empirical knowledge in developing his theory, (b) revising the idea of the circumstances of justice in the Original Position (OP) to include facts about racial realities, and (c) enriching the epistemic resources that discern and denounce racial realities in wide reflective equilibrium. It concludes that option (c) is the most tenable for specifying how considerations of racial realities play a key role in ideal theory to elicit justification for the two principles of justice.

Note that this essay is meant to serve as an introduction to Rawls’s discussion of race, however limited, and the often-thorny debates about the place of race in his ideal theory. Given its restricted purpose, this essay offers a charitable reading that showcases, first, *Rawls’s* thinking about the topic of racial justice and, second, recent attempts in the secondary literature to map the conceptual relation of nonideal facts to the formulation of principles of justice. It disarms Mills’s pointed criticism that Rawls’s ideal-theory methodology is insensitive to racial matters. However, this essay leaves untouched the questions of whether the principles of justice as fairness are sufficient for racial justice or what a Rawlsian theory of racial justice entails.[[1]](#footnote-2) It only seeks to show that anti-racist philosophical and political practices fortify the public grounding of justice as fairness.

#### Rawls’s Case for Excluding Race from a Well-Ordered Society

The distinction between ideal and nonideal theory creates a division of labor for philosophers. Rawls’s two-stage methodology prioritizes ideal theory, under which his contributions mainly fall. Ideal theory identifies the principles of justice for a hypothetical, well-ordered society, in which there is perfect compliance to the requirements of justice. The principles of justice are indispensable for helping nonideal theorists discern injustices; they provide a long-term goal or aim for social justice reforms to achieve. Nonideal theory offers an account of political obligations in the face of avoidable political and institutional failures. Among its chief aims is to diagnose structures of social domination to facilitate the transition to a more just society (Rawls 1999, 137; Laurence 2021; Cf. Shelby 2016, 12). The topics of nonideal theory include civil disobedience, symbolic moral repair for historical injustice and reforming the basic structure to conform to the requirements of justice. Rawls adds that nonideal theory is dependent on ideal theory. “The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems [of injustice]” (Rawls 1999, 6). Rawls affirms that ideal theory articulates the principles and political values essential for promoting gender and racial equality, even if he himself does not develop nonideal theories (2001, 66). While acknowledging that he should have said more about gender and race, he asserts that his relative silence on these matters stems from his methodological commitment to ideal theory. Therefore, he does not regard his omission as an oversight. Let’s examine his arguments for why he excludes race from his representation of a well-ordered society.

To begin, race is not a relevant social position in a well-ordered society. Ideal theory specifies the requirements of justice under “reasonable favorable conditions” (2001, 13). A well-ordered society has “reasonably favorable conditions” in that “the circumstances of justice” obtain. These circumstances delineate “the normal conditions under which human cooperation is both possible and necessary”; and they include the “natural” objective and subjective conditions of social cooperation in a defined geographical region (Rawls 1999, 109). The objective condition assumes neither scarcity nor a superabundance of resources. The subjective conditions are that persons have disparate needs and interests and, as a result, make conflictual claims on limited resources. Rawls adds that people also suffer natural psychological and epistemic limitations; they are neither saints nor incurable egoists (1999, 110). The former would render the idea of justice superfluous, whereas the latter would render it impracticable. Specifying a practical role for the idea of justice would be impossible were philosophers to suppose that it is “unrealistic” to expect people to hamper their self-interest to cooperate for the sake of justice.

A well-ordered society models a fair system of social cooperation. Rawls specifies that people would face three “contingencies” that could generate “inequalities in life-prospects” notwithstanding their good-faith attempts to cooperate:

1. social class of origin: the class into which people are born and develop before the age of reason.
2. native endowments (as opposed to their realized endowments); and their opportunities

to develop these endowments as affected by their social class of origin.

1. good or ill fortune, or good or bad luck, over the course of a complete life (how they’re affected by illness or accident; and say by periods of involuntary unemployment and regional economic decline). (2001, 55)

These contingencies permanently threaten to cause inequalities even in favorable conditions. As a result, the organization of the basic structure must prevent poor life outcomes for those who happen to face them. That neither social class of origin, native endowments, nor fortune should contribute to inequalities is the focus of an ideal theory of justice.

Rawls’s reasons for excluding race as potentially generating inequalities in a well-ordered society is that he proposes that ideal theory answer two critical questions:

first, what contingencies tend to generate troubling inequalities even in a well-ordered society and thus prompt us, along with other considerations, to take the basic structure as the primary subject of justice; and second, how within ideal theory should the least advantaged be specified? (2001, 65)

Race is not the sort of contingency that “tends to generate troubling inequalities” in favorable conditions. A well-ordered society would not contain social positions historically characteristic of racial injustice; rather, all persons regardless of their identity would enjoy equal standing vis-à-vis the basic structure. However, even in a well-ordered society, the basic structure must work to mitigate the “troubling contingences” of social class of origin, unequal native endowments, and fortune. These contingencies would still threaten to contribute to inequalities in the best-case scenario of a good-faith, intergenerational efforts at social cooperation and a strong and stable basic structure. For this reason, Rawls takes the basic structure as the main subject of justice and posits that justice is primarily a virtue that applies to the basic structure. Rawls thus hopes to ensure fair life outcomes in the face of what he understands to be the *ineradicable* contingencies that define the modern human condition, as captured by (1) – (3), damaging people’s life prospects. Striking an optimistic note, he submits that race should not count among the sorts of contingencies that *permanently* threaten to hurt social relations.[[2]](#footnote-3)

Henry Richardson finds an indirect route to object to the exclusion of race and racial injustice from the idea of a well-ordered society. He objects that a well-ordered society would remain vulnerable to the emergence of “fresh” injustices akin to racial injustice (2023, 181). He argues that Rawls offers a “narrowly juridical” conception of the basic structure, in which the mechanisms of structural domination take juridical form. However, dominating power is far more varied, indirect, and complex than Rawls imagines (2023, 188). Rawls’s picture of a well-ordered society implies that racial inequality originates in the unjustorganization of the basic structure. But the principles of justice do not regulate individuals’ actions in interpersonal relations and voluntary civil associations. Because the principles of justice are not meant to govern non-institutionalized norms and individuals’ actions, save in the limited cases of individuals’ holding public office, the legitimate extension of the principles of justice falls short of preventing “elite snobbery” (Richardson 190-97; cf. Kelly 2017, 90). Individuals with superior social capital, such as connections to exclusive social networks and professional opportunities, can acquire illicit power over others. Worse still, background cultural norms will tend to favor elites’ interests and epistemic outlooks. Individual “choices” about “where to live, whom to befriend, and how to treat people in informal contexts” are protected by the principle of equal basic liberty and yet “if widespread would arguably introduce injustice, especially over time” (Matthew 2017, 238). It appears that even a just basic structure would fail to prevent the emergence of new forms of structural domination. In sum, the worry is that Rawls inadvertently implies that the denial of equal basic liberties and fair opportunity define the varied social phenomena of racism to sideline interpersonal exclusionary dynamics among individuals and groups. One needs an argument that demonstrates that racial injustice is delimited by the loss of equal basic liberties and fair opportunity alone, which Rawls does not provide and, in any case, seems wrongheaded.

Contemporary discussion about what constitutes the social phenomena of racial injustice are sophisticated and point to multi-dimensional, interlocking mechanisms of domination. Richardson is no doubt right about this. Yet, textual evidence does not support his narrowly “juridical” reading of the basic structure. Rawls treats the basic structure as a system of social cooperation that encompasses a wide range of extra-legal values and practices that should generate a social union based on reciprocal good will among free and equal civic fellows (1999, 456-64). Emphasizing the underappreciated value of social unity in justice as fairness, some Rawls scholars propose community-oriented, public initiatives to cultivate civic concern for the worst off as part of a well-ordered society’s “friendly” background culture, which should prevent the formation of exclusionary cliques (Schwarzenbach 2007 & 2009). Furthermore, were we to accept Richardson’s legalistic interpretation of the basic structure, it does not follow that it is inappropriate for the state to promote programs of civic education designed to counteract biases and enhance mutual understanding, respect, and civic friendship and other civic virtues among citizens. On the contrary, civic education programs are arguably within the legitimate purview of justice as fairness (Schwarzenbach 2009; Basevich 2019 & 2022). To be sure, it might be the case that appealing to the state to remedy injustice is infeasible or would leave untouched mechanisms of domination that operate apart from the basic structure. Perhaps a two-pronged account of virtue should supplement the idea of a well-ordered society to stress that justice is a virtue of *both* the basic structure *and* of individuals (Peebles 2024).

Another reason why Rawls excludes considerations of race among the sources of pernicious social inequality is that race cannot pick out the worst off *within* a well-ordered society. Race would be an irrelevant social position because the position of the worst off is a non-rigid designator. Inequalities in the distribution of primary social goods determine economic standing. The worst off “is defined […] by reference to […] *an index of primary goods*” (Rawls 2001, 65). This is not to say that Rawls ‘reduces’ considerations of race and gender to class. Rather, as Tommie Shelby notes, the formal requirements of justice prohibit denying persons equal basic liberties and fair opportunities (2004, 1709). These protections enjoy lexical priority over the difference principle in justice as fairness; and their denial to any person or group would be inconsistent with the ideal of equal citizenship. Shelby thus observes “Rawls […] is led to ignore racial characteristics in articulating his theory of justice because he holds that racial inequalities could only be the product of injustice” (2004, 1710). Contra Richardson, he implicitly accepts that securing equal basic liberties and fair opportunity is sufficient to ward off racial injustice. In fact, Rawls posits that once equal basic liberties and fair opportunity are secure, race then becomes an irrelevant point of view because the prospect of racial injustice vanishes:

In a well-ordered society […], the standard relevant positions specified by the primary goods should suffice. We hope that in a well-ordered society under favorable conditions, with the equal basic liberties and fair equality of opportunity secured, gender and race would not specify relevant points of view. (2001, 66)

Once a well-ordered society establishes equal basic liberties and fair opportunities in the basic structure, it distributes primary social goods using the difference principle to benefit the worst off. In this hypothetical scenario, Rawls supposes that there would be no meaningful connection between the categories specified by race and the worst off. Any overlap between these categories would be accidental and an immaterial consideration from the standpoint of justice. Additionally, one need not entertain that race- and gender-based inequalities in the distribution of primary social goods would benefit women and nonwhites. Because race is a non-rigid designator, Rawls rejects that there is a one-to-one correspondence between the worst off and racial identity in a well-ordered society.

The upshot is that Rawls’s exclusion of race from his representation of the circumstances of justice in a well-ordered society does not mean that the principles of justice as fairness are insensitive to racial realities or any other kinds of injustice that the idea of a well-ordered society does not explicitly countenance. Even if ideal theory is meant to mitigate the ineradicable contingencies outlined above, it does not follow that the work of justice is limited to this task alone. Innumerable cases of injustice would fall within the extensional scope of the principles of justice, even if there is little descriptive overlap with Rawls’s description of the circumstances of justice in a well-ordered society and the various circumstances of injustice that abound in reality and are excluded from the picture of a well-ordered society. Furthermore, Rawls describes the principles of justice as general and universal. They are general in that their formulation includes neither rigid designators nor proper names; and they are universal in that they apply to all persons as moral equals (Rawls 1999, 113-14; Cf. Shelby 2004, 1702).[[3]](#footnote-4) In sum, Rawls maintains that a principle of justice should be positioned to redress numerous cases of injustice that ideal theory cannot explicitly countenance because it is beyond its immediate theoretical domain. This is why Rawls argues that justice as fairness has a two-part structure. Ideal theory needs a nonideal theory that incorporates fine-grained descriptions of the social phenomena that ideal principles should redress to actualize the requirements of justice in the circumstances of racial injustice. The role of a nonideal theory is to provide the necessary roadmap.

### Rawls’s and Rawlsian Political Epistemology

That the idea of a well-ordered society is stripped of reference to racial realities does not mean that facts about racial injustice do not play a positive role in ideal-theory construction. On the contrary, while race is absent in Rawls’s account of the circumstances of justice in a well-ordered society, facts about racial injustice inform his justification of the idea of justice as fairness. How best to conceptualize the justificatory power of such facts is a matter of disagreement among Rawls scholars. Below I survey three routes for incorporating consideration of facts about racial injustice to support the justification of the principles of justice as fairness.

#### Rawls’s Anti-Racist Intentions and Beliefs

Some Rawls scholars have tried to determine whether Rawls had antiracist intentions and beliefs when he developed his theory. The assumption here is that Rawls’s *personal* thinking on race sheds insight into whether his ideal theory is responsive to unjust racial realities and justifies a racial justice agenda. Scholars have taken two opposing lines. On the one hand, Mills claims that Rawls’s relative silence on racial justice reflects the limitations of his intellectual history and is “an artefact of racial privilege” (2017, 35). Mills objects that Rawls exemplifies the myopic whiteness of academic philosophy in general (2017, 150). Rawls, the historical figure, did not theorize racial justice because he had the wrong intellectual priorities (Mills 1997, 94; Mills 2013, 15-16). Several philosophers echo Mills’s concern that Rawls and (overwhelmingly white) Rawls scholars never “move on” to race because at bottom “it’s not important to them,” though most do not go so far as to reject justice as fairness as an inadequate normative foundation for racial justice (Matthew 2024, 58-59; Cf. Valentini 2023, 234). On the other hand, Brandon Terry draws on archival research to demonstrate that, on the contrary, Rawls had a deep understanding of, and was personally troubled by, racism (2021, 16). Terry concludes that Rawls was not as silent about racial matters as his published writings suggest. Archival notes detail Rawls’s condemnation of the racial injustice of the Vietnam-War draft in the U.S., which disproportionately conscripted poor people of color. Rawls appears to have voiced antiracist sentiments to his colleagues and students, though he lacked the gumption to defend those sentiments before Harvard University’s administration (Terry 2021, 17-18).

More ambitiously, moving beyond speculative inferences about Rawls’s personal thoughts and feelings, Shelby claims that “there can be no doubt” that race had “a significant influence” on Rawls’s ideal theory, including “on the content of the two principles [of justice]” (2004, 1699). Shelby posits that Rawls’s “thinking on race” had a determinate methodological impact on the idea of justice as fairness. Shelby regards the principles of justice as fairness as somehow “shaped” by Rawls’s extensive knowledge of past and present racial injustices: “The conviction that racial discrimination is unjust helps to shape Rawls’s sense of what is morally relevant and what is morally arbitrary from the standpoint of social justice” (Shelby 2004, 1699). So, too, D.C. Matthew claims that Rawls contemplated a wide variety of historical injustices as part of the “theorist’s pre-writing process of theory development” (2024, 54). They agree with Terry that Rawls regarded his ideal theory as a practical tool for racial justice.

In my estimation, assuming that Terry, Shelby, and Matthew are right, philosophers still need to explain the determinate methodological role that the theorist’s knowledge about racial injustice should play in ideal theory. Rawls’s intellectual history—his silence *and* personal interests—are ultimately immaterial to this task. Were we take for granted that *his* anti-racist sensibilities shaped his theory, the task remains to take the relevant knowledge out of Rawls’s head (assuming, charitably, that it was there in the first place) and specify the role and nature of political knowledge within the conceptual framework of justice as fairness.

Rawls scholars have pursued two options to represent racial realities explicitly at the level of a theory’s justification and development of the principles of justice. One option is to incorporate the relevant knowledge into the Original Position (OP) by thinning the veil of ignorance; another enriches the considered convictions about racial injustice to epistemically reorient reflective equilibrium by drawing on neglected anti-racist philosophical and political practices. Due to space constraints, my discussion stresses that both options showcase that intellectual and political practices oriented around understanding and resisting racial injustice can supporting justifying the principles of justice as fairness. As noted above, I leave aside the difficult implications of my discussion about whether the principles of justice as fairness are sufficient for racial justice. I only show that they require broad-based grounding in epistemic resources that Rawls does not explicitly consider, certifying their racial-justice orientation and public justification.

#### The Original Position

The Original Position (OP) is a hypothetical thought experiment in which representative equal citizens select the principles of justice to govern a well-ordered society from a menu of options. Rawls draws on the OP to demonstrate that justice as fairness is preferable to alternative conceptions of justice and is thus justified. To ensure that representatives reason from an initial choice situation of equality, their knowledge is restricted. They understand that persons have diverse plans of life, basic social and economic theory, and the circumstances of justice. But they do not know their own position in society, the generation into which they’re born, or their society’s developmental conditions. Together with the veil of ignorance, the OP specifies “a point of view from which a fair agreement between *free* and *equal* persons can be reached” (Rawls 2001, 15). The stripped-down epistemic orientation of the OP is designed to ensure that representatives stand as moral equals in an initial choice situation. Without knowledge about particulars, they have no basis for privileging any one group or social position. In response to criticism, Rawls admits race and gender behind the veil of ignorance and stops referring to hypothetical representatives as “heads of households.” Yet it remains important to his reformulation of the OP that it leave intact a thick veil of ignorance. Rawls sets up the OP with an eye towards societal conditions, characterized by the circumstances of justice in a well-ordered society, to support the moral powers of reasonable and rational civic fellows with complete but indeterminable and pluralistic conceptions of the good life. Any information beyond these minimal epistemic constraints is not only irrelevant from the standpoint of justice, but Rawls fears it would import biases that would condone illicit group interests.

As we’ve seen, Rawls’s description of the circumstances of justice in a well-ordered society does not include the experience of racial injustice. It captures the *unavoidable* sources of potential inequalities that could undermine good-faith social cooperation and a just and stable basic structure. But the circumstances of racial injustice reveal *avoidable* sources of social inequalities evident in modern political history. And Rawls countenances that *both* kinds of contingencies—unavoidable inequalities rooted in the circumstances of justice *and* the avoidable inequalities of the modern racialized world*—*constitute together the sorts of social inequalities over which the principles of justice have legitimate domain and mustextend. Some Rawls scholars favor a thinner veil to represent *explicitly* the circumstances of racial injustice as a salient and urgent problem from the public standpoint of justice.[[4]](#footnote-5) So much so, they assert that explicit consideration of nonideal facts should practically motivate and epistemically reorient normative political inquiry.

For example, Erin Kelly builds on Okin and Mills to favor a thinner veil. Kelly foregrounds the avoidable and historically contingent racial inequalities “to emphasize that the social significance of collective wrongdoing and its legacy shape the problems of justice that the resources of political philosophy, including the rich resources of Rawls’s theory, can and should help us to think about” (2023, 170). Kelly reformulates the idea of the circumstances of justice to capture racial realities, past and present. Resisting Rawls’s reliance on what she calls “a pre-political state of nature” to describe the modern human condition using the stripped-down version of the circumstances of justice, she includes in the OP “contingent facts about the history of the society to be governed[.] The parties to Rawls’s original position,” she argues, “should consider, in their deliberations, their society’s history of war and other forms of politically organized violence and domination, at least when that history has resulted in enduring forms of social inequality that call for rectification, on grounds of justice” (2023, 169).

Unlike Mills (but like Okin) Kelly does not abandon the two principles of justice as fairness.[[5]](#footnote-6) She instead “urges that we view Rawls’s two principles themselves as principles of corrective justice, properly interpreted as sensitive to the harm structures of domination have inflicted on the social bases of self-respect, as well as the material interests, of oppressed people” (2023, 179). In her estimation, the principles of justice as fairness acquire a novel, justificatory power once theorists admit facts about racial injustice into the OP. Namely, consideration of facts about racial injustice illuminates the practical value of the principles of justice as fairness to achieve moral repair in the aftermath of collective wrongdoing (Kelly 2023, 170-71). The purpose of a public conception of justice encompasses a reformative function in that the novel value—or “the good”—of justice includes its capacity to “reconcile […] groups whose relationship has not been a relationship of reciprocity and mutual trust”(Kelly 2023, 177). Kelly thus challenges philosophers to reinterpret the purpose, if not the content, of the principles of justice as fairness. This reformative reorientation should inform how philosophers interpret the justification of the principles of justice as fairness.[[6]](#footnote-7) For the best overall conception of justice is one whose principles can reconcile groups in the aftermath of historical injustice.

Kelly’s proposal is creative and ingenious. Unfortunately, it also unravels Rawls’s defense of the OP to adduce justification for justice as fairness. Recall that the idea of the circumstances of justice has a specific role in Rawls’s conception of a well-ordered society. It captures the inescapable objective and subjective conditions of human cooperation in a definite geographical region with the aim of supporting representatives’ selection of a suitable conception of justice for a democratic society of free and equal civic fellows. Representatives in the OP are not supposed to reason in the light of facts about any given world. Rather, the idea of the circumstances of justice serves the important normative function of situating hypothetical representatives in a symmetrical relation in the OP. In conjunction with ignorance about historically contingent facts, the OP is designed to elicit a sense of justice that works to eliminate the influence of bias and self-interest on deliberation. Representatives must select the principles of justice in an initial choice situation that models the *equality* of moral persons. Persons’ moral powers are “the *basis of equality,* the features of human beings in virtue of which they are to be treated in accordance with the principles of justice” (Rawls 1999, 441; Freeman 2007, 159). Hypothetical representatives adopt these epistemic constraints in connection to a conception of citizenship that assumes that persons possessing two moral powers, that of being reasonable and rational, which any viable theory of justice must honor. Rawls thus retains a “Kantian” constructivist commitment by stripping the explicit knowledge accessible to representatives in the OP to ensure that the procedure of selection respects the moral equality of persons, conceived as autonomous co-citizens with diverse plans for a good life.

Once we explicitly admit historically contingent facts about injustice into consideration for representatives, it is impossible to regard the OP as an initial choice situation of equality. On the contrary, facts about racial injustice illuminate deep and systematic asymmetries of social power that form ascriptive social hierarchies. Perhaps normative theorizing ought to begin from these nonideal premises, but it is difficult to see how the OP would continue to serve the specific function of providing justification for a public conception of justice that free and equal persons ought to accept. Such a radical, new departure point would require an innovative and complex alternative justification for the idea of justice as fairness.[[7]](#footnote-8) Perhaps Kelly’s thought here is that the selection of candidate conceptions of justice should primarily reflect how well a candidate conception repairs damaged social relations, as if the task of theorizing justice should be seen as fundamentally a project of effectuating reconciliation in the aftermath of historical wrongdoing.[[8]](#footnote-9) In any case, Rawls wants to “keep things simple” (1999, 108). Representatives are to assume a fixed picture of societal circumstances in weighing conceptions of justice in the OP. For this reason: “it turns out to be desirable to characterize the original position so that parties are to choose principles that hold unconditionally whatever the circumstances” (1999, 108-9). *After* representatives select justice as fairness, the thick veil can lift for representatives to consider facts that imbue the principles of justice with a reformative reorientation. This suggests that *explicit* consideration of facts about racial injustice do not belong in the OP. But do they belong in ideal theory at all?

#### Reflective Equilibrium

There is another route for Rawls scholars to take to capture the reality of racial injustice within ideal theory, one that I believe is most promising because it showcases how history and political practice inform philosophical inquiry: the method of reflective equilibrium. Rawls describes reflective equilibrium as “another side to justifying a particular description of the original position. This is to see if the principles which would be chosen match our considered convictions of justice or extend them in an acceptable way” (Rawls 1999, 17). In fact, the OP is “embedded” in the “broader” method of reflective equilibrium (Valentini 2023, 224). One uses it to align one’s judgments and general principles to achieve a coherent position, first, by ordering one’s convictions in a narrow equilibrium and, second, by seeking out alternative perspectives, philosophical traditions, and principles to enter a wide reflective equilibrium. In the process of critical self-reflection, one should give up convictions if they do not cohere with principles one accepts, or if one’s convictions are not supportable by the balance of public values and perspectives. “Crucially, to do reflective equilibrium *well* is to do it in broad conversation with others. People from dif­ferent backgrounds and with different life experiences have different insights and liabilities when it comes to moral discernment” (Schouten 2024, 186). Reflective equilibrium requires deliberative agents, ideally, to go outside themselves to consider the experiences, evaluative perspectives, and intellectual traditions of other people and their take on what’s wrong with world. The method of reflective equilibrium provides an alternative description of an initial standpoint of equality characteristic of the OP by modeling “an impartial judgment not likely to be distorted by an excessive attention to our own interests,” but it must proceed with consistent and empathetic engagement with the situated perspective and experiences of other people (Rawls 1999, 17-18).

Rawls also admits substantive moral content that makes “free use of the facts” in his treatment of the considered convictions that enter reflective equilibrium:

There is a definite if limited class of facts against which conjectured principles can be checked, namely, our considered judgments in reflective equilibrium. […] The analysis of moral concepts and the a priori, however traditionally understood, is too slender a basis. Moral theory must be free to use contingent assumptions and general facts as it pleases. There is no other way to give an account of our considered judgments in reflective equilibrium. (1999, 44)

The empirical information embedded in considered judgments encompasses knowledge of the world and the past. Considered judgments convey insights about historical injustice and social inequality that help establish a standard for distinguishing right from wrong from a first-personal deliberative perspective and establish “provisional fixed points which we presume any conception of justice must fit” (1999, 18). While potentially revisable, these are judgments “we never expect to withdraw, as when Lincoln says: ‘If slavery is not wrong, nothing is wrong’” (Rawls 2001, 29). Considered judgments are thus not mere descriptions of the world; the empirical facts they contain should contribute to principled insights about best how to organize a just society.

Considered judgments are not mere mirror reflections of what any given actual person thinks. Rawls affirms that his theory rests on substantive evaluative commitments that a people already hold *or* would hold upon due reflection, emphasizing that the outcome of reflective equilibrium can result in moral transformation. Reflective equilibrium meets people where they are, “working from our existing considered convictions” (Freeman 2007, 35). Yet, he also stresses that some convictions form the epistemic heritage of a pluralistic constitutional democracy. He assumes that history attests to the advance of collective moral insights. By and large, the historical practices that have shaped modern political communities have not been gross moral failures, racked by one moral disaster after another. Upon their critical self-reflection, people can come to see their political, philosophical, and historical traditions in a new light to motivate their support for the idea of justice as fairness. By aligning in wide reflective equilibrium, considered judgements add prima facie justification to the idea of justice as fairness in that the principles of justice as fairness should, ideally, complement established interpretative norms and social practices. As an exercise of practical reason, the method shows that Rawls’s ideal theory of justice is historically mediated and context sensitive (Schwarzenbach 1991; Valentini 2023, 226).[[9]](#footnote-10) The view that Rawls’s ideal theory unfolds in a vacuum devoid of critical reflection on history and shared experience is simply a mistake that both critics and defenders of Rawls often make.

Consider that Rawls draws heavily on the history of anti-racist intellectual and political practices, affirming the injustice of slavery and racial discrimination, to illustrate the method of reflective equilibrium at work. The convictions in which a people *should* have “confidence” include “that religious intolerance and racial discrimination are unjust” (1999, 17). Shelby thus observes: “Consideration of the historical and current realities of racial injustice […] have an indirect role to play in ideal theory through the process of coming to reflective equilibrium.” Shelby continues:

We must inevitably rely on our firmest pre-theoretic moral convictions about particular cases of injustice when evaluating our theoretical conclusions, treating these convictions as fallible and “provisional fixed points” for moral reflection, not as self-evident moral foundations. So if the principles chosen from the original position didn’t condemn, for example, slavery, racial segregation, or racial discrimination in employment, then we could be confident that these principles aren’t sound and that we need to revise the conditions of the initial choice situation. Ultimately, we want our general conclusions to cohere with our deepest convictions about particular cases so that philosophical theory and concrete moral judgments mutually support each other. The realities of past racial domination are not therefore to be ignored in ideal theorizing, though we may abstract away from them for purposes of using the contractarian apparatus to justify general principles. (2013, 151)

Shelby captures that inquirers’ firm, pre-theoretic convictions should express an anti-racist moral sensibility, resting on people’s familiarity with assorted key facts about racial realities and resistance, past and present. He surmises that in the U.S. “our” epistemic heritage rests on deep anti-racist convictions (2004, 1704). While early American history was once a hotbed of racist ideology, luckily, “here and now,” considered judgments are sufficiently robust to stand as “provisional fixed points” that add justificatory weight to justice as fairness, one that clarifies its implicit racial-justice agenda.

Shelby’s appeal to “settled,” pre-theoretical convictions that people “currently” have is question-begging: it passes the buck in an epistemically unsatisfying fashion and presupposes what we’ve yet to establish (Shelby 2004, 1704; 2013, 151): the method of reflective equilibrium must reject social arrangements that use race to determine groups’ access to social goods and opportunities; and it must shore up the anti-racist epistemic and moral resources for inquirers to achieve such an informed position. Those resources, however, might not be widely available in officially sanctioned public spaces. Shelby plays down the fact that consideration and resistance to racial injustice is not well represented in most political communities’ moral and epistemic traditions. Save in the case of the abolition of slavery, anti-racism is certainly not a vantage point that philosophers should posit as a fixed normative reference point in the context of political communities with brutal histories of racial terror and discrimination that lasted for at least a century after the abolition of the trans-Atlantic slave trade. In sum, the worry is that the experience of racial injustice is not well represented in dominant philosophical and political traditions, though it should be. If one appeals to considered convictions to justify the sensitivity to racial injustice to support the idea of justice as fairness, then one needs to defend a deliberative model that foregrounds the relevant anti-racist considerations, even if they are not “deeply” held by many and primarily manifest in marginalized black radical and Africana philosophical and intellectual traditions and political practices. In other words, to make good on Rawls’s conceit that his ideal theory is sensitive to racial realities, then philosophers need to take the next step to show that the cultivation of anti-racist sensibilities rest on neglected philosophical and political traditions, without which the idea of justice as fairness would lack grounding in its extensions to a racial-justice agenda. For it is a stretch to presuppose a coherent and sensible view is in the offing in the formal public sphere about a matter that remains deeply contentious.

I believe the method of reflective equilibrium can be clarified and expanded. Philosophers can provisionally foreground which considered convictions and the philosophical and political practices commonly associated with them ought to enjoy an epistemically privileged status to orient deliberation, at least provisionally, in the early stages of reflective equilibrium. Rawls gives ample room to accommodate novel moral and epistemic insights that can reorient reflective equilibrium. Philosophers can highlight the relevant considerations about the circumstances of *racial* injustice, which should help inquirers avoid “mistakes” and “common excuses” (1999, 47-48). The upshot is that “settled” and “considered” convictions could be developed *with the aid of* “alternative [hermeneutic] horizons of interpretation, judgement, and critique opened up by, for example, black radicalism” (Terry 2021, 21; cf. Alcoff 2024). These latter evaluative practices surely count among the standpoints which merit consultation in wide reflective equilibrium. Furthermore, in their absence, it is impossible to achieve considered convictions that would amount to a meaningful anti-racist moral sensibility. Philosophers can thus “radicalize and extend into new domains” the method of reflective equilibrium by drawing on anti-racist philosophical traditions and political practices, as well as a myriad progressive and neglected feminist and anti-racist social justice movements. Sibyl Schwarzenbach makes the point nicely:

[T]he vast repertoire of particular, considered moral convictions hitherto relegated to the “private,” nonpolitical, and noneconomic spheres—whether the diaries of the housewife, the manifestos of the slave, or the rantings of the abolitionist—must now be drawn into the original data pool from which a society-wide reflective equilibrium at least begins. (2009, 231-32).

Consider an epistemic provision, specified by philosophers, that discloses *ascriptive social hierarchies* andfunctions as a provisional “fixed point” in reflective equilibrium. Figuring prominently in critical race theory and social justice movements, the idea of ascriptive social hierarchies captures that a person’s social identity often imposes an unfair distribution of the benefits and burdens of social cooperation. Ascriptions of inferiority *cause* women and people of color to lose social goods. Ascriptive social hierarchies distribute rights and resources on an ideological basis and are not mere byproducts of unequal distributive schemes. Rather social devaluation *qua* group membership leads to poor life outcomes (Anderson 2012).

The epistemic provision holds that eliminating ascriptive social hierarchies is an urgent and practical task that specifies *an indispensable vantage point* from which to assess the merits of competing conceptions of justice and, ultimately, support the justification of the idea of justice as fairness in wide reflective equilibrium. The purpose of representing facts about racial injustice within reflective equilibrium is to showcase that the principles of justice as fairness must work to dismantle ascriptive social hierarchies as a “test” of their merit. For the most suitable conception of justice should be one that is inconsistent with ascriptive social hierarchies of any kind, whether based in race or any other social categories. And this is precisely the concern of black radical and Africana philosophical traditions and contemporary social justice movements. Far from being insensitive to the reality of racial domination, the idea of justice as fairness must find novel *grounding* in these marginalized perspectives. For philosophers can provide a compelling reinterpretation of the moral and epistemic resources among marginalized groups to show why, taken together, they offer a critical perspective that can enhance the public standpoint of justice and add to its justification.

My proposal retains the benefits of Kelly’s proposal without the interpretative difficulties presented by revising Rawls’s account of the circumstances of justice in the OP. With the above provision in mind, philosophers can appreciate that the practical value of principles of justice involves their capacity to repair damaged social bonds in the aftermath of injustice. Unlike deliberation within the OP, wide reflective equilibrium offers an alternative description of the *socio-historically situated* conditions of the contractual parties without tarnishing a thick veil. Though the OP is embedded in the method of reflective equilibrium, empirically laden judgments about racial injustice do not reappear in the OP as facts for explicit consideration before representative agents choosing a conception of justice from a menu of options. Rather, they facilitate inquirers’ critical self-clarification and capture that the public standpoint of justice is historically shaped by philosophical and political traditions confronting racial injustice.

### Conclusion

This chapter canvasses how Rawls and Rawls scholars have treated the conceptual relation of ideal theory to facts about racial realities. My argument proceeded in two stages. First, I defended Rawls’s exclusion of facts about racial injustice from his picture of a well-ordered society. I then considered what role, if any, such facts must play in ideal theory construction, focusing on the methods of the Original Positions and Reflective Equilibrium. I have defended an epistemic provision that tentatively reorients reflective equilibrium to alternative philosophical traditions and political practices to stress an anti-racist moral sensibility that rests on the rejection of ascriptive social hierarchies. In this way, I hope to have showed that Rawls’s ideal theory is not a vacuous “ideal ideal” that is insensitive to racial realities. Rather, at its core, the method of reflective equilibrium is epistemically reflexive and must accommodate novel considered judgments and alternative viewpoints. In fact, inquirers should ‘test’ theories to demonstrate their incompatibility with ascriptive social hierarchies. Modern systems of oppression are complex and ever evolving. Though my presentation had focused on racial realities, the epistemic provision might also flag gender, sexual orientation, immigration status, and national origin. Across any and all circumstances persons should never face undue social burdens arbitrarily imposed on them because of their group membership (Cf. Shiffrin 2004). The task of dismantling ascriptive social hierarchies stands as the defining practical concern of a theory of social justice.

References:

Alcoff, Linda M. “The roots (and routes) of the epistemology of ignorance,” *Critical Review of*

*International Social and Political Philosophy*27.1 (2024): 9-28.

Anderson, Elizabeth. “Equality,” in David Estlund (ed). *Oxford Handbook on Political Philosophy*.

New York: Oxford University Press, 2012.

Basevich, Elvira. “Self-Respect and Self-Segregation: A Duboisian Challenge to Kant and Rawls,”

*Social Theory & Practice* 48.3 (2022): 403-27.

Basevich, Elvira. “W.E.B. Du Bois’s Critique of American Democracy in the Jim Crow Era: On

the Limitations of Rawls and Honneth,” *Journal of Political Philosophy* 27.3 (2019): 318-340.

Du Bois, W.E.B. *The Souls of Black Folk*. New York: Oxford University Press, 2007.

Freeman, Samuel. *Rawls*. New York: Routledge, 2007.

Haslanger, Sally. “Methods of Social Critique,” *Publications of the Austrian Ludwig Wittgenstein*

*Society* 28 (2020):139-56.

Kelly, Erin I. “The Circumstances of Justice,” Paul Weithman (ed.) *Rawls’s A Theory of Justice at*

*50*. New York: Cambridge University Press, 2023. pp. 169–80.

Kelly, Erin I. “The Historical Injustice Problem for Political Liberalism,” *Ethics* 128.1 (2017): 75-

94.

Laurence, Ben. *Agents of Change: Political Philosophy in Practice*. Cambridge: Harvard

University Press, 2021.

Matthew, D.C. “Rawls and racial justice,” *Politics, Philosophy, & Economics* 16.3 (2017) 235-

58.

Matthew, D.C. “Rawls’s ideal theory: A clarification and defense,” *Res Publica* 25 (2019): 553-

70.

Matthew, D.C. “Mills’ *The Racial Contract* and Ideal Theory,” *Critical Review of International*

*Social and Political Philosophy*27.1 (2024): 47-61.

Mills, Charles W. *Black Rights/White Wrongs: The Critique of Racial Liberalism*. New York:

Oxford University Press, 2017.

Mills, Charles W. “Retrieving Rawls for Racial Justice?: A Critique of Tommie Shelby,” *Critical*

*Philosophy of Race* 1.1 (2013): 1-27.

Mills, Charles W. and Carol Pateman. *Contract & Domination*. Cambridge: Polity Press, 2007.

Okin, Susan. *Justice, Gender and the Family*. New York: Basic Books, 1989.

Peebles, Ian S. “Toward a virtue-based account of racism,” *Philosophical Studies* 181.10 (2024),

2499-523.

Rawls, John. *Political Liberalism*. New York: Columbia University Press, 2005.

Rawls, John. *Justice as Fairness: A Restatement*. Erin Kelly (ed.). Cambridge: Harvard

University Press, 2001.

Rawls, John. *A Theory of Justice*. 2nd Edition. Cambridge: Harvard University Press, 1999.

Richardson, Henry. “Why Rawls’s Ideal Theory Leaves the Well-Ordered Society Vulnerable to

Structural Oppression,” in Paul Weithman (ed.) *Rawls’s A Theory of Justice at 50*. New

York: Cambridge University Press, 2023. pp. 181-98.

Schouten, Regina. *The Anatomy of Justice*. New York: Oxford University Press, 2024.

Schwarzenbach, Sibyl. “Rawls, Hegel, and Communitarianism,” *Political Theory* 19(4): 539-71.

1991.

Schwarzenbach, Sibyl. *On Civic Friendship Including Women in the State*. New York: Columbia

University Press, 2009.

Shelby, Tommie. “Race, Reparations, and Justice as Fairness,” in Paul Weithman (ed.) *Rawls’s A*

*Theory of Justice at 50*. New York: Cambridge University Press, 2023. pp. 199–219.

Shelby, Tommie. *Dark Ghettoes: Injustice, Dissent, and Reform*. Cambridge: Harvard University

Press, 2016.

Shelby, Tommie. “Racial Realities and Corrective Justice: A Reply to Charles Mills,” *Critical*

*Philosophy of Race* 1.1 (2013): 145-62.

Shelby, Tommie. “Race and Ethnicity, Race and Social Justice: Rawlsian Considerations,”

*Fordham Law Review* 72.5 (2004): 1697-1714.

Shiffrin, Seana V. “Race, Labor, and the Fair Equality of Opportunity Principle,” *Fordham Law*

*Review* 72.5 (2004): 1643-75.

Terry, Brandon M. “Conscription and the Color Line: Rawls, Race, and Vietnam,” *Modern*

*Intellectual History* (2021): 1-24.

Valentini, Laura. “On the Role of the Original Position in Rawls’s Theory,” in Paul Weithman (ed.)

*Rawls’s A Theory of Justice at 50*. New York: Cambridge University Press, 2023. pp. 220-

36.

1. For an excellent discussion of this second issue, see D.C. Matthew (2017). Matthew argues that Rawls’s ideal theory can do “reformative” work to mitigate the harms of past racial injustices. Given the reformative scope of the principles of justice, “at least some of the need to draw on nonideal theory is *obviated*” (2017, 244-45). [↑](#footnote-ref-2)
2. Susan Moller Okin objects that the gendered division of reproductive labor does permanently threaten to cause unjustifiable gender-based inequalities. To counteract this tendency, she reconstructs the OP to include “the knowledge that women have been and continue to be the less advantaged sex in a great number of respects” (1989, 102-3). [↑](#footnote-ref-3)
3. Rawls also stipulates that principles of justice are public in that everyone accepts the same conception of justice, the basic structure is just and these two facts are public knowledge. [↑](#footnote-ref-4)
4. Susan Moller Okin includes nonideal facts about sexist injustice in the OP to show that the principles of justice as fairness must support the abolition of the gendered division of labor and render reproductive labor in the home as a central problem of justice (Okin 1989, 102-3). [↑](#footnote-ref-5)
5. She also rejects the ideal/nonideal theory distinction. However, she does not modify the *content* of the principles of justice as fairness. Mills dramatically *weakens* the requirements of justice as fairness. He endorses equal basic liberties and favors formal—over fair—equality of opportunity. (Mills 2007, 121; Cf. 2013, 21; 2017, 179-80). [↑](#footnote-ref-6)
6. Cf. (Matthew 2017, 244-45). [↑](#footnote-ref-7)
7. Mills revises the OP as a procedure for sorting through the *least unjust* competing conceptions with full knowledge of injustice (2017, 213). Though he doesn’t specify the outcome of this decision-making procedure, it appears that it wouldn’t converge with the principles of justice as fairness. [↑](#footnote-ref-8)
8. Admittedly, Rawls is sympathetic to this idea (2001: 3-4, 115) [↑](#footnote-ref-9)
9. Rawls observes that the “historical origin of political liberalism” lies in the Wars of Religion, the Reformation, and “its aftermath” (2005, xxiv). [↑](#footnote-ref-10)