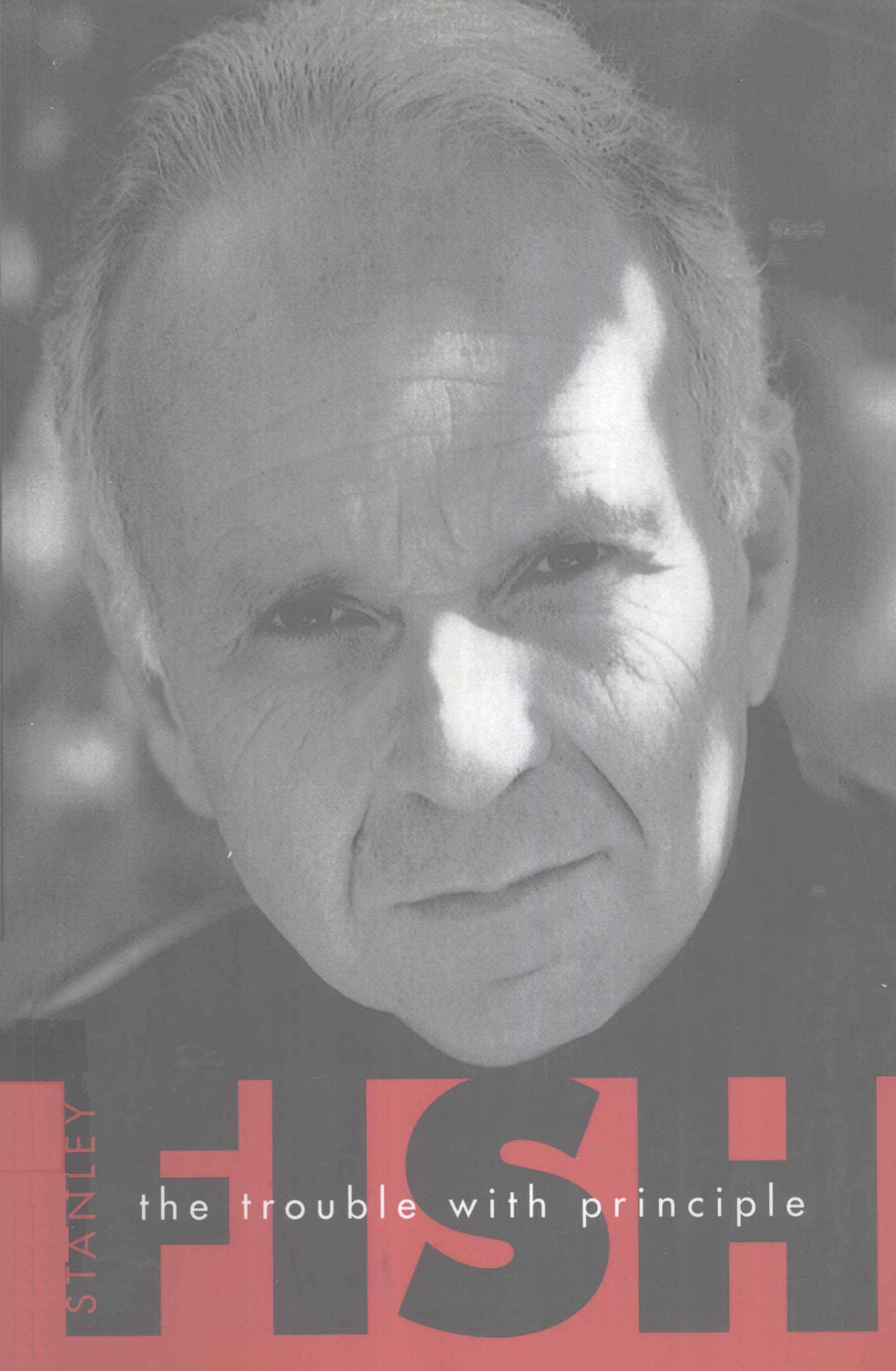

STANLEY

FISH

the trouble with principle

THE TROUBLE WITH PRINCIPLE

Stanley Fish

Harvard University Press

Cambridge, Massachusetts

London, England

Printed in the United States of America
Second printing, 2001
First Harvard University Press paperback edition, 2001

Library of Congress Cataloging-in-Publication Data
Fish, Stanley Eugene.
The trouble with principle / Stanley Fish.
p. cm.
Includes bibliographical references and index.
ISBN 0–674–91012–5 (alk. paper)
ISBN 0–674–00534–1 (pbk.)
1. Principle (Philosophy). 2. Political science—Philosophy.
3. Law and politics. I. Title.
B105.P7F57 1999
323'.01—dc21 99–35759

CONTENTS

Prologue

While I was writing the chapters of this book, a scene from Sam Peckinpah's classic western *The Wild Bunch* was never far from my mind. The wild bunch is an outlaw gang led by two grizzled veterans played to a career-performance turn by William Holden and Ernest Borgnine. One evening the two are sitting around discussing an old comrade who has gone over to the other side and now rides at the head of the band of railroad detectives pursuing them. The Borgnine character is incensed and can't understand why their old friend doesn't abandon the pursuit and come home to where he really belongs. You have to remember, the Holden character says, he gave his word to the railroad. So what? is the response; it's not giving your word that's important, it's who you give your word to.

I read the scene as a profound and concise analysis of the great divide in political theory. On the one side is the man of principle for whom a formal contract must be kept irrespective of the moral status of the other party; when you give your word, you give your word and that's it. On the other side is the man who varies his obligations according to the moral worth of the persons he encounters; some people have a call on your integrity, others don't, and the important thing is to determine at every moment which is which.

There is, I think, no doubt about which of these two visions is today the more generally approved. The Holden character speaks in the

accents of Enlightenment liberalism; what he says is in accord with maxims many of us have long since internalized: "A man's word is his bond." "Ours is a government of laws, not men." "You can't justify the means by the end." "Respect for your fellow man must be extended to all and not selectively." Each of these maxims urges us to enter a perspective wider than that formed by our local affiliations and partisan goals; each gestures toward a morality more capacious than the morality of our tribe, or association, or profession, or religion; each invites us to inhabit what the legal philosopher Ronald Dworkin calls "the forum of principle," the forum in which our allegiances are not to persons or to wished-for outcomes but to abstract norms that neither respect nor disrespect particular persons and are indifferent to outcomes.

Not that there has never been a strong argument on the other side. The Borgnine character is not alone in his sentiments, and among those who would support him in the exchange (although they would be an odd couple) is John Milton. Milton and his characters are always saying things like "You are not worthy to be convinced" (the Lady to Comus in the mask of that name), or "You don't owe any loyalty to a king who is not acting like one" (Milton to his countrymen in *The Tenure of Kings and Magistrates*), or "Everyone should be allowed to speak and publish, except of course Catholics" (Milton to the Parliament in the *Areopagitica*). When Satan describes himself to the angel Gabriel as a "faithful leader" (*Paradise Lost,* IV, 933), the angel immediately replies, "Faithful to whom? To thy rebellious crew?/ Army of fiends?" (953–954). Like the Borgnine character, Gabriel refuses a notion of fidelity that is indifferent as to its object; some are deserving of your faith, some others are not, and to maintain loyalty merely because you have once pledged it is to mistake an abstraction for an object of worship and to default on your responsibility first to determine what (or who) is good and true and then to follow it.

Let me say at the outset that I am with Borgnine and Milton and against an adherence to principle. The trouble with principle is, first, that it does not exist, and, second, that nowadays many bad things are done in its name. On the surface, this is a paradox: how can something that doesn't exist have consequences? The answer is to be found in the claim made for principle, or at least for the kind of principle—usually called neutral principle—favored by liberal theorists. The claim is that abstractions like fairness, impartiality, mutual respect, and reasonableness can be defined in ways not hostage to any partisan agenda. The

importance of the claim is that if it can be made good, these and other abstractions can serve as norms or benchmarks in relation to which policies favoring no one and respecting everyone can be identified and implemented. The problem is that any attempt to define one of these abstractions—to give it content—will always and necessarily proceed from the vantage point of some currently unexamined assumptions about the way life is or should be, and it is those assumptions, contestable in fact but at the moment not contested or even acknowledged, that will really be generating the conclusions that are supposedly being generated by the logic of principle.

If, for example, I say "Let's be fair," you won't know what I mean unless I've specified the background conditions in relation to which fairness has an operational sense. Would it be fair to distribute goods equally irrespective of the accomplishments of those who receive them, or would it be fair to reward each according to his efforts? Is it fair to admit persons to college solely on the basis of test scores and grades, or is it fair to take into account an applicant's history, including whatever history he or she may have of poverty and disadvantage? Such questions sit at the center of long-standing political, economic, and social debates, and these debates will not be furthered by the simple invocation of fairness, because at some level the debate is about what fairness (or neutrality or impartiality) really is.

Moreover, even if the concept of fairness is filled in, you are by no means out of the woods. Perhaps I think it's fair when everyone has a chance to have his say, but you think it's fair when everyone who is qualified to speak has a chance to have his say. Any challenge to either of our stipulations will result in further stipulations (decisions are best when every citizen has participated in making them, or decisions are best when the forum is limited to those who are educated or who have been elected, or who are ordained or who are not ordained), which in turn will be subject to challenge because they are substantive.

This last word—substantive—is the key, for it is supposedly the virtue of neutral principles to be free of substantive commitments; it is within the space afforded by neutral principles, or so we are told, that substantive agendas can make their case without prior advantage or disadvantage, with the result that the best argument (best is never defined; or, if it is, it is vulnerable to the same questions I have put to fairness) will win. But what the example of fairness—and you could substitute impartiality or neutrality or any other formal universal and it would

turn out the same—shows is that there are no neutral principles, only so-called principles that are already informed by the substantive content to which they are rhetorically opposed. And even if you could come up with a principle that was genuinely neutral—a notion of fairness unattached to any preferred goal or vision of life—it would be unhelpful because it would be empty (that, after all, is the requirement); invoking it would point you in no particular direction, would not tell you where to go or what to do. A real neutral principle, even if it were available, wouldn't get you anywhere in particular because it would get you anywhere at all.[1]

Curiously enough, this is what makes neutral principles so useful politically and rhetorically and gives them the capacity to do bad things. It is because they don't have the constraining power claimed for them (they neither rule out nor mandate anything) and yet have the *name* of constraints (people think that when you invoke fairness you call for something determinate and determinable) that neutral principles can make an argument look as though it has a support higher or deeper than the support provided by its own substantive thrust. Indeed, the vocabulary of neutral principle can be used to disguise substance so that it appears to be the inevitable and nonengineered product of an impersonal logic. There are many ways to play this game, but in all its versions the basic move is to turn historically saturated situations into situations fully detached from any specific historical circumstance and then conclude that a proposed policy either follows from this carefully emptied context or is barred by it.

An example (and it is an example of doing a bad thing) is the majority argument in *Plessy v. Ferguson* (1896), a case that turned on the constitutionality of a Louisiana law requiring railroad companies to provide separate but equal accommodations for whites and blacks.[2] The plaintiffs challenging the law argued that it violated the Thirteenth and Fourteenth amendments. The Supreme Court ruled that there was no such violation and chided the plaintiffs for assuming "that the enforced separation of the two races stamps the colored race with a badge of inferiority": "If this be so it is not by reason of anything found in the act but solely because the colored race chooses to put that construction on it."[3] Here the phrase that does the crucial work is "in the act," which means "in the act *itself*," that is, the act as it is, apart from the history that gave rise to it or the intentions of those who performed it. Abstracted from

this background information, "the act" is indeed without the significance imputed to it by the plaintiffs; it is without any significance at all. And therefore it makes perfect sense to say, as the Court has said earlier in its opinion, that laws requiring the separation of the two races "do not necessarily imply the inferiority of either" (50). This is too weak; such laws, understood without reference to the conditions of their production, imply *nothing* necessarily but can be made to imply anything.

This is what happens when this or any other "act" is detached from the history that renders it intelligible; it becomes unreadable, or (it is the same thing) it becomes readable in any direction you like. But if you reinsert it into the historical context of its production, implications are easily and nonarbitrarily readable, as they were for Justice Harlan when he said in his dissent, "Everyone knows that the statute in question had its origin in the purpose, not so much to exclude white persons from railroad cars occupied by blacks, as to exclude colored people from coaches occupied by or assigned to white persons" (52). Harlan alerts us to the nature of the work "in the act" did for the majority; it removed the purpose—the motive, aimed-at outcome, political agenda—informing the legislation and rendered it entirely formal, empty of substantive historical content. It might as well have been the separation of molecules or quarks or jelly beans that the Court was speaking of.

Harlan also makes clear that the Court was doing its work under the cover of a neutral principle (although he doesn't call it that), the principle of nondiscrimination: "It was said in argument that the statute of Louisiana does not discriminate against either race, but prescribes a rule applicable alike to white and colored citizens" (52). As Harlan sees, this is disingenuous at best. It is true that mixing in the railroad cars is denied to both races and thus, on this point, the rule does not discriminate against either; but it is also true that the rule against mixing flows from the design of one of the races to keep the other down. That design is certainly discriminatory, and the present act is certainly its extension, but the majority is able to obscure (perhaps from itself) that fact by defining nondiscrimination formally, as a matter of principle, and one that is applied to parties assumed to have no histories or antecedent purposes, parties that might just as well have been identified as X and Y rather than as black and white. Once that abstracting gesture has been performed, once the origins of the Louisiana statute have been forgotten, the Court can declare with a straight face that it discriminates against no one (if only because there is now no one—no particular one—in sight).

A neutral principle that facilitates the forgetting of history is repeating the forgetting that allowed it to emerge *as* a neutral principle. Neutral principles, if they are to deserve the name, must be presented as if they came first, as if they were there before history, even if the inhabitants of history were slow to recognize them. A neutral principle, in short, can have a historical habitation but not a historical cause. Accordingly, the question one asks of it is analytic ("What is its essence?") rather than genealogical ("Where did it come from?"). But once the genealogical question is put and the principle is given a biography, the idea of regarding it as neutral—as without reference to substantive imperatives—will seem less compelling.

Ask, for example, where the principle of nondiscrimination comes from. It comes not from some a priori moral map—it did not come down with the Ten Commandments—but from a set of historical circumstances to which it was a response. Men did things to other men for a long time and without moral anxiety. At some point (and rather slowly) moral anxiety arose, and with it a new name for old activities: discrimination. Because the new name was understood to be a negative judgment, it generated an initiative of reform ("Let's do away with this unhappy practice"); and as the proponents of reform warmed to their task, they began to think of it as the Lord's work and as the extension of an absolute and pre-existing imperative, the imperative of nondiscrimination. Once this last move is made, it becomes possible (and tempting) to think of nondiscrimination in formal terms, as a principle that mandates policies favoring no group no matter who the group is or what it has done or what has been done to it. When that thought has taken hold, the principle has become a neutral principle to which you can be true only by abstracting away from the historical concerns that gave rise to it.

In his dissent, Justice Harlan is trying to recall his brethren to those concerns and that history, asking them in effect to remember why the Thirteenth and Fourteenth amendments were enacted in the first place. The standard story we tell one another is that while Harlan failed in 1896, he was vindicated in 1954 when the separate-but-equal doctrine was overturned in *Brown v. Board of Education.* But while that particular doctrine has been repudiated, the logic underlying it—the logic of neutral principles—has not been, and today the courts regularly issue rulings the *Plessy* majority would have recognized and approved. The vehicle of *Plessy*'s resurrection has been the notion of reverse discrimination, the assertion that any action tinged with race-consciousness is

equivalent to any other action tinged with race-consciousness, an assertion that makes sense only if historical differences are dissolved in the solvent of a leveling abstraction. Armed with this neutral principle, this device for erasing the difference between oppression and the amelioration of oppression, the courts have undone the gains of affirmative action and issued rulings in which the Voting Rights Act, passed in order to enhance the voting power of blacks, is violated by any policy that is true to the act's purpose. Indeed, it would not be too much to say that the result of these rulings has been to declare the Voting Rights Act in violation of itself.

This is what I meant when I said at the beginning that many bad things are now being done in the name of neutral principles, and I hope it is clear by now that it is no paradox to say that bad things are being done by something which doesn't exist. Indeed, it is crucial that neutral principles not exist if they are to perform the function I have described, the function of facilitating the efforts of partisan agents to attach an honorific vocabulary to their agendas. For the effort to succeed, the vocabulary (of "fairness," "merit," "neutrality," "impartiality," "mutual respect," and so on) must be empty, have no traction or bite of its own, and thus be an unoccupied vessel waiting to be filled by whoever gets to it first or with the most persuasive force.

But while there is a strong relationship between the emptiness or nonexistence of neutral principles and the work that they do (again, the emptiness provides the space for the work), there is no relationship at all between the emptiness of neutral principles and the political direction of that work. I have labeled the things I see being done with neutral principles "bad" because they involve outcomes I neither desire nor approve. They are not "bad" simply because they were generated by the vocabulary of neutral principles, for that vocabulary has also generated outcomes I favor, especially in the areas of civil rights and the expansion of opportunities for women in the workplace and on the athletic field. The fact that the game of neutral principles is really a political game—the object of which is to package your agenda in a vocabulary everyone, or almost everyone, honors—is itself neutral and tells you nothing about how the game will be played in a particular instance. The truth, as I take it to be, that neutral principles, insofar as they are anything, are the very opposite of neutral, and are filled with substance, won't tell you what substance they are filled with or whether or not you will like it.

The fact that someone is invoking neutral principles will give you no clue as to where he is likely to come out until he actually arrives there and reveals his substantive positions.

And whatever *my* substantive positions are, they do not follow from the fact that I believe neutral principles to be the empty vehicles of partisan manipulation. Reading the previous pages will have given you no clue as to my views on any vexed question. It would have been perfectly consistent with my analysis of *Plessy* had I announced at the end of it that I was in agreement with the majority. My interest in the analysis was in the way the Court managed to turn a statute manifestly discriminatory into an exemplary instance of nondiscrimination. Any indignation I may have registered should be traced to my dislike of the outcome and not to my disapproval of the strategy. The passion I display when debunking the normative claims of neutral principle ideologues is unrelated to the passion I might display when arguing for affirmative action or for minority-enhancing redistricting. To be sure, there might be a contingent relation in a given instance if the outcome I dislike was brought about in part by neutral-principle rhetoric; I might then attack the rhetoric as part of my attack on what it was used to do. But I might turn around tomorrow and use the same rhetoric in the service of a cause I believed in. Nor would there be anything inconsistent or hypocritical about such behavior. The grounding consideration in both instances (whether I was attacking neutral-principle rhetoric or employing it) would be my convictions and commitments; the means used to advance them would be secondary, and it would be no part of my morality to be consistent in my handling of those means.

I know that some of my readers will think that I have revealed myself in the preceding sentence to be one of those horrible persons who profess a morality of ends rather than means. But that has been my stance from the beginning. The argument that neutral principles are either empty or filled with the substance they claim to hold at arm's length is an argument for the impossibility (despite many claims and hopes) of disentangling oneself from substantive agendas and therefore an argument that ends-based reasoning cannot be avoided. It is not, however, an argument *for* ends-based reasoning (arguing for something you can't avoid would be an odd thing to do), and in making it I am as usual offering no recommendation (you can't coherently recommend an inevitability), just pointing out, for the umpteenth time, that when all is said and done there is nowhere to go except to the goals and desires that

already possess you, and nothing to do but try as hard as you can to implement them in the world.

No doubt this will be heard by some as the unhappy and suspect announcement that everything is politics. This would be a strong and damning point only if politics meant the implementation of naked preferences. Naked preferences, however, are a part of the neutral-principle picture of the world, where it is possible (and desirable) to distinguish between principle and substance. In that picture, naked preferences are the danger and neutral principles the bulwark erected against them. But there are no naked preferences for the same reason that there are no neutral principles: principle and substance come always mixed. Principle and its vocabulary of fairness, equality, and so on are already informed by substantive preferences (were they not, they would be incapable of giving direction), and preferences are always preferences in relation to some notion of the good; they are never naked. In fact, preferences (except for trivial cases like a preference for vanilla ice cream over chocolate) *are* principles (or at least principled)—not principles of the neutral kind but principles of the only kind there really are, strong moral intuitions as to how the world should go combined with a resolve to be faithful to them.

For liberal theorists, however, such principles are not enough because they are insufficiently general. But the demand for a general principle—for a principle to which one might turn when moral values clash and there is a need for adjudication—is once again the demand for a neutral principle. A principle would be general if it could be applied invariably to variable contexts and situations; a general principle resists appeals to the particular, resists statements like "You must take into consideration that he's had a hard life," or "You should bear in mind the practical effects of your decision," or "You should remember how important this project is to the goal of racial peace." You arrive at the identification of a general principle by putting any candidate to the test proposed by Herbert Wechsler (who introduced the term "neutral principles" to the legal community): "Would I reach the same result if the substantive interests were otherwise?"[4] If the answer is no, if the principle you are applying generates different outcomes when one or the other of the parties changes color, or the organization at the center of a case is the NAACP rather than the Ku Klux Klan, or the publication under attack is pornography rather than *The New York Times,* then it is not really general, is not really a principle, but is instead a policy

weighted in the direction of a particular race or class of groups or type of expression.

The generality of a principle, then, is measured by the extent of what it ignores, and in human terms—the terms in relation to which most moments in our lives are led—a general principle ignores almost everything, not only race, gender, class, religious affiliation, ethnic identity, sexual orientation (the usual list), but accomplishments, failures, value to society, moral worth. These negative requirements for reaching the plateau of general principle are perfectly met by John Rawls's "original position," a position one occupies by willfully putting on a "veil of ignorance"—ignorance of "features relating to social position, native endowment, and historical accident, as well as to the contents of persons' determinate conceptions of the good" (79), or, as Rawls calls them, "comprehensive doctrines."[5] Comprehensive doctrines are an embarrassment to the hopes of generality because there are too many of them (two would be too many). This is a "serious problem," says Rawls, because a modern society "is characterized not simply by a pluralism of comprehensive religious, philosophical, and moral doctrines but by a pluralism of incompatible yet reasonable comprehensive doctrines," no one of which "is affirmed by citizens generally" (xii). Under such conditions, Rawls asks, what does one do? And the answer he gives is: "abstract from and not be . . . affected by the contingencies of the social world" (23); that is, "find some point of view, removed from and not distorted by the particular features of [any] all-encompassing background framework" (25).

Such a view is what Thomas Nagel calls the view from nowhere, and in Rawls's argument it is also the view from no one, from no particular one, from a one that is general. Achieving that view, he claims, "serves as a means of public reflection and self-clarification" (26), but it would be more accurate to name the result "self-evacuation," for the imaginative act of entering the original position involves a bracketing and (for the time of sojourn) forgetting of every affiliation and association that makes you what you are.

The trick is to regard social, political, and institutional investments as cosmetic. One sees how it is done when Rawls describes reasoning in the original position as proceeding "in accordance with the enumerated restrictions on information" (27). The restrictions are the sum of what you are not allowed to know under the veil of ignorance—everything

from name, rank, and serial number to matters of gender, class, and race to memberships in churches and political parties to "various native endowments such as strength and intelligence" (25). By referring to these as restrictions on *information,* Rawls makes it clear that in his view the characteristics they remove from inspection are not essential to the person, who is what he is with or without these identifying marks of merely social relations: he is an agent with a capacity to imagine a condition of justice and a vision of the good (81); and it is this capacity, rather than any realization it happens to have, that defines him. Those who have this capacity, even if they realize it in different ways, are the same; and a person who realizes it differently at different times in his life is also the same: "For example, when citizens convert from one religion to another, or no longer affirm an established religious faith, they do not cease to be, for the questions of political justice, the same" (30). "For the purposes of public life, Saul of Tarsus and Paul the Apostle are the same person" (32n).

It would appear that by declaring Saul and Paul to be the same only in the public realm, Rawls gives the experience of conversion its due, but in fact he trivializes it by limiting its effects to the quarantined realm of the private, where religious commitment takes its place alongside commitment to the Elks Club or the New York Yankees or country music. One imagines that Paul would have a different view of the matter, and that if you told him it was all right to believe anything he liked (even that someone walked on water and rose from the dead) as long as he left his belief at home and didn't allow it to influence his actions in public, he might smile at you and say things like "You cannot serve two masters."

It is no accident that the question of religion surfaces often in Rawls's book, for religion is the very type of the substantive views that are to be banished, or at least set to the side, in a world ordered by neutral principles. Religion is the chief stumbling block in the way of the liberal dream of a public life cleared of the perturbations that arise when fundamental beliefs (liberal theorists rename them "opinions" on the way to marginalizing them) come into conflict. Rawls acknowledges this at the very beginning: "The most intractable struggles, political liberalism assumes, are confessedly for the sake of the highest things: for religion, for philosophical views of the world, and for different moral conceptions of the good" (4). He proposes to deal with this intractability as John Locke proposed to deal with it three hundred years ago, by removing issues of religion, philosophy, and morality from the public

agenda. But if these are really the "highest things," a public realm purged of them would be diminished and thin and you will have bought your peace at the price of substance (a price liberal theorists are willing and eager to pay); and, on the other hand, if you think that public life emptied of religious, philosophical, and moral urgencies is robust and substantial, you don't believe that they are the highest things and you call them that only as a gesture on the way to dismissing them.

Political liberalism stakes its project on the possibility of at once being fair to religion, giving it its due, and preventing its concerns (the concerns of moral conviction in general) from interfering with the operations of the public sphere. But it can manage this only by turning the highest things into the most ephemeral things (higher in the sense of "airy") and by making the operations of the public sphere entirely procedural, with no more content than the content of traffic signals. In fact, neither of these strategies can be realized; the religious impulse will refuse to be confined (that's what makes it what it is), and matters of procedure will never be purely so, will always be touched by and touch on matters of substance. Rawls wonders whether, given the deep oppositions that have always divided men along religious, philosophical, and moral lines, "just cooperation among free and equal citizens is possible at all" (4). It isn't.

What is possible is cooperation achieved through the give and take of substantive agendas as they vie for the right to be supreme over this or that part of the public landscape. In the course of such struggles, alliances will be formed and for a time at least conflict of a deep kind will be kept at bay. Alliances, however, are temporary, conflict is always just around the corner (Hobbes was right), and when it erupts, all the muted claims of "comprehensive doctrines" will be reasserted, until, for largely pragmatic reasons, those claims are again softened and replaced (temporarily) by the conciliatory words of another vocabulary, perhaps the vocabulary of political liberalism. As a genuine model for the behavior of either persons or nations, as something you could actually follow and apply, political liberalism is hopeless. Like all projects based, supposedly, on neutral principles, it is either empty (you can't get from its abstractions to the nitty gritty of any actual real-life situation) or filled with an agenda it cannot acknowledge lest it be revealed as the limiting and exclusionary mechanism it surely is. That is, the project either doesn't exist or it exists on a level less general and considerably lower than the level of its polemic.