1989

Authority and Value: Reflections on Raz's Morality of Freedom

Donald H. Regan

University of Michigan Law School, donregan@umich.edu

Follow this and additional works at: http://repository.law.umich.edu/articles

Part of the Jurisprudence Commons, and the Law and Philosophy Commons

Recommended Citation


This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.
Joseph Raz’s *The Morality of Freedom* is full of subtle, original, and thought provoking arguments. It also manifests abundantly Raz’s philosophical good sense and sensitivity to the complexities of the moral life. These are reasons enough to class it with the handful of genuinely important books whose appearance in the last two decades has constituted a renaissance in political philosophy.

But in my opinion, Raz has another, and even stronger claim on our attention: He comes closer to the truth about political morality than anyone has for nearly a century. (Possibly much longer, but we need not attempt to decide exactly who last did as well as Raz.)

The most important reason for Raz’s getting nearer the truth than others have of late is that Raz is (or seems to me to be) a consequentialist. He does not say he is a consequentialist. Indeed, he would surely claim, and rightly, not to be the sort of theorist most people think of as consequentialist. But neither does Raz ever deny he is a consequentialist, at least not with any conviction. I should say he describes his explicit position perfectly (and in all its uncertainty) when he says he wants to show “practical thought is either non-consequentialist or consequentialist in an attenuated sense only . . . .”

† © 1989 Donald H. Regan. All Rights Reserved.

* Professor of Law and Professor of Philosophy, University of Michigan. B.A. 1963, Harvard University; LL.B. 1966, University of Virginia; M. Phil. (Economics) 1968, Oxford University; Ph.D. (Philosophy) 1980, University of Michigan. I would like to thank members of the Legal Theory Workshops at McGill University and the University of Virginia; and I thank yet again that tireless reader of his colleagues’ drafts, Fred Schauer.

2. Id. at 265-66.
The source of Raz's ambivalence is clear. There is no generally accepted definition of just what makes a consequentialist, and Raz explicitly rejects some of the doctrines traditionally associated with consequentialist thinking. For a start, he rejects "strict consequentialism," the doctrine that the only reasons for or against an action are the consequences of the action (or of its nonperformance), where "consequences" do not include the action itself.3 Strict consequentialists must believe either that actions have no intrinsic value or that their intrinsic value does not count in deciding whether to perform them. Raz is right to reject strict consequentialism. (I doubt whether any serious consequentialist has ever been a strict consequentialist, but that is another matter.) Raz also rejects "transparency," the doctrine that states of affairs are intrinsically valuable only to the extent they are good for some agent or other, and that goodness for an agent is determined by that agent's subjective evaluation.4 Again, Raz is right to reject transparency. In rejecting it, he separates himself from a large part of the consequentialist tradition, but not, for example, from G.E. Moore.

Finally, Raz rejects "comparability," the doctrine that all reasons for action are comparable in respect of their strength, and that they can be ranked with regard to strength, singly or in combination.5 Here I am inclined to part company with Raz (that is to say, I believe in comparability), but it seems to me that one who rejects comparability can still be a consequentialist. The reader may wonder how it is possible to maximize, as consequentialism requires, over incomparable values. We cannot straightforwardly maximize in the fullest sense; but we can engage in what I shall call "quasi-maximization," of which Pareto-optimization is a well-known instance. We first evaluate each possible state of affairs with respect to each of the incomparable value dimensions. We then say that one state of affairs dominates another if and only if it is at least as good on every value dimension, and better on at least one. The principle of quasi-maximization requires us to produce some state of affairs which is not dominated by any other feasible state of affairs. We then say that one state of affairs dominates another if and only if it is at least as good on every value dimension, and better on at least one. The principle of quasi-maximization requires us to produce some state of affairs which is not dominated by any other feasible state of affairs. Quasi-maximization will leave us a good deal more choice among eligible (that is, non-dominated) states of affairs than ordinary maximization does. But quasi-maximization does impose significant constraints on choice; and it is the natural generalization of ordinary maximization in the face of incomparability, just as Pareto-optimization is the natural surrogate for utility-

3. See id. at 268-70.
4. See id. at 269-71, 294-99.
5. See id. at 321-66.
maximization if we have doubts about interpersonal comparisons of utility. When I discuss the issue of incomparability, I shall suggest that Raz sometimes wants to squeeze out of his theory more definite conclusions than quasi-maximization really allows; I doubt that Raz hews consistently to his belief in incomparability. But that is beside the present point, which is that one who quasi-maximizes in the face of incomparable values acts in the spirit of consequentialism.

Fundamentally, a consequentialist is someone who believes that morality, whether individual or political, is about promoting the occurrence in the world of whatever has intrinsic value. (Incidentally, belief in the essential unity of individual and political morality is one hallmark of the consequentialist tradition which Raz does not include in his list of consequentialist doctrines, but which he explicitly shares.) Raz fits this definition of a consequentialist. "If reason guides choices it does so by pointing to the value of different courses of action and of their consequences." Raz fits this definition of a consequentialist. "If reason guides choices it does so by pointing to the value of different courses of action and of their consequences."7 "Since autonomy is morally valuable there is reason for everyone to make himself and everyone else autonomous."8 "Morality is . . . concerned with the advancement of the well-being of individuals."9 "[G]overnments should promote the moral quality of the life of those whose lives and actions they can affect."10 And so on.

Starting from the other end, if Raz is not a consequentialist, what is he? He unequivocally rejects the notion that morality is founded on rights, whether of autonomy or of equality.11 Rights aside, Raz countenances the possibility of "intrinsic duties," but it turns out that such duties require acts that, by their expressive nature, either help to constitute some intrinsically valuable relationship, such as friendship, or manifest respect for such values as the value of art.12 Even here, claims about what is intrinsically valuable are at the bottom of it all. So, I shall say that Raz is a consequentialist, perhaps to some extent malgré lui. If I regarded "consequentialist" as a term of opprobrium, I would be reluctant to fix it on anyone over his reservations. But since I regard it as an imputation of mere good sense, I shall not shrink.

A second reason why Raz comes closer to the truth than others have done is that despite being a consequentialist, he rejects the views most

6. See id. at 4.
7. Id. at 363.
8. Id. at 407.
9. Id. at 267.
10. Id. at 415.
11. Id. at 163-263.
12. See id. at 210-13.
common among consequentialists concerning what has intrinsic value or what we should promote. As part of his rejection of the doctrine of transparency, he rejects the claim that pleasure is the good, as well as the claim that we should promote the satisfaction of desires as such. (I believe he would equally reject the notion that what counts is "informed desire" or "rationally criticized desire"). The ultimate good is (or most often appears to be) individual well-being. The well-being of individuals is constituted in large measure by the success of their relationships and projects, but only of those relationships and projects which are actually valuable. The value of relationships and projects is not guaranteed simply by their being chosen by some agent. This is the direction in which consequentialism must go.

I shall say more presently about Raz's views on well-being and on the comparability of values. For now, let me add just one more general observation on Raz's ambiguous relationship to consequentialism. Much of the flavor, and much of the richness, of The Morality of Freedom stems from a specific, pervasive tension. On the one hand, Raz appreciates the appeal of consequentialist reasoning and is unwilling flatly to reject the claim that promoting the good is what morality is all about. On the other hand, he wants obligations to be a bit more substantial, and relationships like friendship to be a bit "thicker," and our commitments to our projects to be a bit more existentially significant, than he is sure is really consistent with consequentialist thinking.

I think there is a good reason and a bad reason for this tension in Raz's thought. The good reason is that he is embarked on a very difficult project which has received too little attention—the project of identifying what we really do think is good, once we agree it is not pleasure or the satisfaction of desire. Once we try to say just what is valuable in friendship, or just what should be our attitude and behavior to great art, or just how it is that certain actions have intrinsic value and disvalue, we find that these questions are much more difficult than they may have appeared. Any sensitive philosopher will find himself to some degree inclined to hold views about what is valuable that seem hard to square with the idea that we ought always to be promoting all possible values impartially, even if this requirement is merely "in principle." In the end, I think there is no insuperable difficulty here for a carefully conceived neutralist consequentialism; but I confess that at present that is as much an article of faith as an elaborated philosophical position.

13. See id. at 298-99.
As to the bad reason for the tension I have described in Raz, I think he wants obligations to be a bit more substantial, friendship to be a bit "thicker," and commitments to particular projects to be a bit more existentially significant than they really are.\textsuperscript{15} I shall explain in due course.

In this essay, I cannot possibly discuss all parts of Raz's rich book. Let me therefore summarize the contents quickly, so that the reader may be aware of much interesting material I shall pass over without comment in the main discussion.

An initial chapter sets out "The Problem of Political Freedom."\textsuperscript{16} Raz identifies himself as a liberal (which is not, as Raz means it, inconsistent with being a consequentialist and even a perfectionist) and he says his object is "a view of the ideal of individual freedom and its role in politics."\textsuperscript{17} Specifically, Raz wants to rebut the "revisionist challenge" to the value of freedom, which takes two forms. One form of the challenge concedes that freedom is valuable instrumentally, but denies that it is valuable intrinsically. The other form concedes that certain definite freedoms—freedom of conscience, freedom to participate in government, economic freedoms—are intrinsically valuable, but denies that freedom in general is. Raz wants to argue that freedom in general is intrinsically valuable (which does not entail that any activity becomes valuable by being freely chosen).

Raz's answer to the revisionist challenge, as he presents it, depends on certain claims he makes concerning the value of autonomy, claims which I think he does not establish.\textsuperscript{18} So in a sense I think the book fails in its central stated goal. But if one is really concerned to answer the revisionist challenge, there is material enough in Raz to show that the revisionists are wrong, even if the proper account of the value of freedom leaves it sounding a bit less important than Raz would have it.

After the initial chapter, the book is in five parts. Part I, "The Bounds of Authority," I shall discuss at some length.\textsuperscript{19} Part II is entitled "Anti-Perfectionism."\textsuperscript{20} Raz of course turns out to be an anti-anti-perfectionist. The argument of this part, aimed especially at John Rawls and Robert Nozick, is piecemeal and does not lend itself to summary. It

\textsuperscript{15} See infra text accompanying notes 38-95, 208-15 (regarding obligation); see infra text accompanying notes 157-70 (regarding friendship); see infra text accompanying notes 185-92 (regarding commitments).
\textsuperscript{16} J. Raz, supra note 1, at 1-19.
\textsuperscript{17} J. Raz, supra note 1, at 2.
\textsuperscript{18} See infra text accompanying notes 171-207.
\textsuperscript{19} See infra text accompanying notes 30-98.
\textsuperscript{20} J. Raz, supra note 1, at 107-62.
does, however, add up to a convincing refutation of the claim that government should not act on any particular conception of the good.

In Part III, "Individualistic Freedom: Liberty and Rights," Raz advances novel arguments to prove that neither rights to autonomy nor rights to equality can be the foundation of political morality. The argument against fundamental rights to autonomy, which appears in Chapter Eight, can be summarized as follows: (1) genuine autonomy requires the existence of a wide range of social practices which define activities and relationships (such as being a doctor, being a concert violinist, marriage, friendship, and so on); (2) the existence of any such social practice is a collective good; (3) individuals do not have rights to the existence of such collective goods; therefore, (4) individuals do not have rights to the conditions of autonomy. (Remember that this is a summary. Raz argues for each of the premises with considerable subtlety.)

The argument against a significant fundamental right to equality, which appears in Chapter Nine, can also be summarized: (1) equality cannot be all our political morality cares about, because we only care about equality in the distribution of things that are independently recognized as good; (2) all the plausible candidates for fundamental goods are (a) "satiable" and (b) "diminishing"—which is to say (a) that there comes a point, as an individual acquires more of the good, at which she has a full complement (for example, a person can be fully happy, in such a way that further successes in her projects do not make her happier), and (b) that the closer someone is to having a full complement of the good, the less urgent it seems to move her closer still, as compared to improving the lot of someone else who is further from satiation; it therefore transpires that (3) there is simply no significant role for a principle of equality to play—considerations of the relative urgency of each individual's claim on satiable, diminishing goods are enough by themselves to require an essentially egalitarian distribution of any fundamental goods.

Parts II and III together constitute Raz's critique of nearly all the liberal theorizing of the last few decades. I am in complete agreement with Raz's conclusions; therefore, I shall not undertake any full-dress

21. Id. at 163-263.
22. Id. at 193-216.
23. At one point Raz seems to deny that he relies on precisely the argument given in the text. I am not persuaded by the denial. Unquestionably, Raz has other things to say against rights to autonomy, things that may in the end be more persuasive than this argument. Still, I think this is a fair summary of what is offered most prominently as the direct argument against rights to autonomy. For more on this point see infra notes 127-34 and accompanying text.
24. J. RAZ, supra note 1, at 217-44.
discussion of these parts of Raz's book. I certainly commend them to the reader's attention. I shall have a few things to say about the matter of equality in my discussion of well-being, and about Raz's argument against rights to autonomy in my discussion of social forms.

Part IV, which I shall discuss at some length, brings together Raz's discussions of consequentialism, the nature of well-being, and the incommensurability of values, under the general heading of "Society and Value." Finally, Part V, "Freedom and Politics," which I shall comment on more selectively, develops Raz's notion of autonomy and argues for a version of John Stuart Mill's harm principle as a general constraint on governmental coercion or manipulation.

I. AUTHORITY

To set the stage, let me describe my position vis-à-vis Raz in the broadest terms. Most of Raz's discussion of authority (what I shall call the "general" part) makes no reference to consent or to respect for law. So far as the general part of the discussion goes, I agree completely with Raz's analysis of the structure of the relationship between an authority and the persons subject to that authority. That is to say, I agree with Raz's "three theses" (at least, as I would interpret them). I disagree, however, with Raz's interpretation of the relationship. Specifically, I think it is a mistake to characterize the relationship in terms such as "obedience," "obligation," or "legitimacy." With regard to the special part of Raz's discussion, involving consent and respect for law, the situation is much the same. I agree with Raz that an attitude of identification with one's reasonably just society is intrinsically valuable, and that acts of compliance with law may be intrinsically valuable as expressions of such an attitude. But once again, I do not think these facts establish anything we should refer to as an "obligation" to obey the law.

A. THE STRUCTURE OF AUTHORITY: RAZ'S THREE THESES

Raz's ultimate interest is in political authority, the authority of law. But he begins with an extended discussion of practical authority in general. "Practical" authorities are, roughly speaking, authorities we look to in the process of deciding what to do. They are agents or agencies

25. See infra text accompanying notes 106-07.
26. See infra text accompanying notes 117-34.
27. J. RAZ, supra note 1, at 265-366; see infra text accompanying notes 99-170.
29. See infra text accompanying notes 171-207.
whose utterances are in some sense reasons for action, as opposed to mere reasons for belief, such as are provided by scholarly authorities or the like. (I shall say more about the distinction between reasons for action and reasons for belief presently.) Hereafter, when I say "authority" I shall mean practical authority.

Raz's view of the nature and justification of (practical) authority in general can be summarized in three theses. First, the "dependence thesis": "[A]ll authoritative directives should be based on reasons which already independently apply to the subjects [i.e., the addressees] of the directives and are relevant to their action in the circumstances covered by the directive." As Raz points out, the dependence thesis does not entail that authorities may act only to promote the interests of those subject to them, since the subjects themselves may have reasons to act which are not just a matter of promoting their own interests. Nor does it entail that the authority can do nothing but announce the resultant force of various pre-existing reasons. Each citizen has reason to contribute to the national defense or the relief of hunger in his society, but even so, he has no reason to pay income tax until there are tax laws in place. So tax laws add something to the obligation to contribute to common goods. But they are consistent with the dependence thesis because they should be based on the pre-existing reasons to contribute (or other similar reasons).

The next thesis about authority is the "normal justification thesis": "[T]he normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly." Briefly, Jones is an authority for me (perhaps with regard to a certain range of decisions) if I will do a better job of doing what I have reason to do (with regard to that range of decisions) by doing what Jones tells me to do than by following my own direct judgment.

Finally, there is the "preemptive thesis": "[T]he fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing

30. J. Raz, supra note 1, at 47.
31. See id. at 48.
32. Id. at 53.
what to do, but should exclude and take the place of some of them."

This thesis applies only to legitimate authorities, by which I take it Raz means persons whose claims to authority are validated by the normal justification thesis or otherwise. And even persons who are normally legitimate authorities need not be obeyed in all circumstances. For example, if we know the authority did not take into account all relevant reasons in making its decision (because important new evidence has appeared since an arbitrator decided a case or because an emergency has arisen, presenting quite new circumstances), or if we know the authority took into account irrelevant reasons and so acted arbitrarily, then we can ignore the authority.

So, the point of the preemption thesis is not that authority is absolute, but only that it is special. The reason for action that is constituted by an authoritative directive is not just one reason among others, to be weighed along with all the rest. Rather it is a reason which summarizes and represents certain other reasons (the ones the authority is authoritative about and has presumptively considered). Because the authoritative directive summarizes and represents these other reasons, it displaces them as grounds for the subject's action; it makes it inappropriate for the subject of the authority to base his action on these reasons directly, or, therefore, to base his action on his own judgment regarding their force. The authority's judgment controls.

Now, subject to some minor qualifications or reformulations which for the most part need not concern us, I think Raz's three theses about authority are correct. But the relationship they characterize between authority and subject, or at least the minimal relationship between authority and subject that satisfies them, is much "thinner" than Raz's language suggests. Specifically, Raz's three theses are satisfied by what I shall refer to as the "indicator-rule conception of authority."

B. INDICATOR-RULES AND THE INDICATOR-RULE CONCEPTION OF AUTHORITY

Before I explain what I mean by the indicator-rule conception of authority, I must explain what I mean by an indicator-rule. To locate

33. Id. at 46.
34. See id.
35. See id. at 42.
36. See id. at 46.
37. See id. at 42, 46.
the idea quickly and crudely, an indicator-rule is something more substantial than a rule of thumb, but less substantial than a rule simpliciter (even after we take into account that most rules simpliciter impose only prima facie obligation). Actually, what I now refer to as an indicator-rule is what I have thought of for years as a rule of thumb. But it has gradually become clear to me that most people mean rather less by "rule of thumb" than I have meant; so I have given up that phrase. The main point, in any event, is to be clear about how the sort of rule in question works. I am not aware of anyone who has clearly described, with all its features, what I shall now describe under the label "indicator-rule."

To begin with, an indicator-rule is a guide-to-action which one adopts to deal with cases where one has incomplete information, or to conserve resources spent on decisionmaking, or to counteract one's tendencies to partiality, or the like. If this suggests a "mere" rule of thumb, that is fine, for the moment. For most people, the easiest way to understand an indicator-rule is probably to start with their conception of a rule of thumb and then to add on certain features (or alternatively to bring to consciousness features of their inchoate understanding, as the case may be).

The first important feature of indicator-rules specifically is that they need not be infallible. Some things we call "rules of thumb" are infallible, and it may be part of some people's conception of rules of thumb that they must be infallible. For example, an infallible arithmetic rule for determining whether a large number is divisible by nine is to determine whether the sum of its digits is divisible by nine. This rule of thumb for determining divisibility by nine can save us calculation time, and if it is properly applied, it can never lead us astray. But if we were confronted with some problem for which there was no rule which would save us time without ever leading us astray, it could well be worth using a rule which saved us time, and which usually got the right answer (or which usually came close to the right answer, in cases where "close" was meaningful and good enough), even though it did sometimes lead us astray. So, a good indicator-rule (whether adopted to save time, or to deal with incomplete information, or to counteract bias) may sometimes get the wrong answer. Even though it sometimes gets to the wrong answer, it may be the best guide-to-action that we have, all things considered.

The next point concerns the "bindingness" of indicator-rules. People sometimes say that rules of thumb are not binding, or that we

(properly) feel no compunction when we violate rules of thumb.\textsuperscript{39} Whatever may be the case with rules of thumb (and of course, what is the case with them depends on just what people refer to by "rules of thumb"), indicator-rules can be "binding" in a significant sense.

We need another example. Let us assume for purposes of argument (what I take to be true) that sex between unmarried persons is not invariably immoral, but that it is to be avoided if either party's participation is not fully voluntary. Under these circumstances, an excellent indicator-rule for a university faculty member is not to have sexual relations with any of his (or her) students. (Many universities are now adopting formal rules along these lines, but my point is that this is a good indicator-rule for a faculty member even in the absence of any university rule or policy.) This indicator-rule is not infallible. It is not impossible that a student should have sex fully voluntarily with her (or his) professor, and so the indicator-rule in question might lead one to avoid a non-harmful, perhaps even a genuinely valuable, relationship. But it is very unlikely that the student's participation is fully voluntary, even if the professor does nothing that could possibly be thought of as actively coercive. So, "Never have sex with your students" is a good, albeit fallible, indicator-rule.

This indicator-rule can be thought of as "binding" in a very important sense: A professor may be bound as a conscientious decisionmaker to adopt the indicator-rule and to act on it. To see this, we proceed in two stages. First, we notice that the indicator-rule is based on values which the professor is bound to respect regardless of his own desires or preferences. People may sometimes depreciate rules of thumb because they think of them as adopted merely to facilitate pursuit of the agent's contingent aims. But that is not the case with the indicator-rule now before us. (I would not necessarily deny the title "indicator-rule" to a rule which was adopted only to help in achieving contingent aims. But I would not claim such an indicator-rule was binding in any sense. My present point is that some indicator-rules promote noncontingent, obligatory ends; and these are the ones that may be binding in the sense I am now concerned with. A further point: It is of course a contingent matter whether any particular agent actually concerns himself with those non-contingent, obligatory ends he ought to be concerned with. And in that sense it is a contingent matter whether any particular agent adopts some


HeinOnline -- 62 S. Cal. L. Rev. 1005 1988-89
indicator-rule which he should adopt. But this is true of the agent's relation to any kind of moral value or moral rule at all. It cannot be relied on to deny the bindingness of indicator-rules in particular.)

Now, having noticed that the indicator-rule against sex with one's students is based on values the professor is bound to respect, we can go on to observe that this indicator-rule may well be the best guide-to-action available to the professor for deciding whether to have sex with his students (because of the cost of information and calculation, the possibility of bias, and so on). If it is the best guide-to-action available, then he is bound as a conscientious decisionmaker to adopt the indicator-rule and to act on it. Furthermore, he should feel compunction if he fails, without adequate reason, to adopt it or to act on it.

I say the professor is "bound as a conscientious decisionmaker." My first thought was to say "bound in conscience," but that suggests a mode of being bound which is somewhat too weak. A distinction is frequently made between subjective and objective moral obligation. But in fact, there is a spectrum here. The most "subjective" sort of obligation is just the obligation to follow one's best moral judgment, however defective that moral judgment may be. This is the obligation to "follow one's conscience." At the other end, the most "objective" sort of obligation is the obligation to do what is actually the right thing to do—what one would know to be right if one were fully informed, fully sensitive to all moral considerations, and fully rational in one's thinking. In between, there is an obligation to adopt the best decision procedure available to one for making moral choices. The best decision procedure is a response to incomplete information, limited rationality, and so on, so it may not identify infallibly one's actual, objective obligation in the strongest sense of "objective." But equally, adopting the best available decision procedure requires something more specific and less subjective than just "following one's conscience." The point is, of course, that one ought to adopt and act on a particular indicator-rule if it is part of the best available decision procedure for dealing with the moral problems one faces. In that sense, one is bound as a conscientious decisionmaker to adopt and to act on certain indicator-rules, namely, the best indicator-rules one is aware of for promoting such actual values as one is required to promote.

So far, I have been explaining how it is that indicator-rules are more substantial—more binding—than most people seem to think rules of thumb are. It remains to explain how indicator-rules are less substantial than rules simpliciter. The important point here is that even though conscientious agents are bound to adopt and to act on certain indicator-
rules, those indicator-rules still have no *intrinsic* importance. We can see this in two ways.

First of all, we can notice that the features of the world that the indicator-rules focus on are not the features that have fundamental moral significance. In our example, the indicator-rule focuses on the fact that some person is a student taught by another person. But the indicator-rule does not reflect a belief that there is some intrinsic inappropriateness in sexual relations between teacher and student. (Someone might think that, but that has not been our working assumption.) Rather the indicator-rule reflects an assumption that if there is a student-teacher relationship, the student’s participation in the sexual relationship is unlikely to be fully voluntary. This is a very plausible empirical generalization, but it is an empirical generalization nonetheless. It need not and does not reflect a belief that the existence of a student-teacher relationship matters *fundamentally*, or in the ultimate analysis, to the question whether a sexual relationship between two parties is proper.

We can make the point clearer by contrasting the present indicator-rule with the rule that one should keep one’s promises. On the standard view (which it might be well to mention I do not share), the fact that one has made a promise is ordinarily a fact of *fundamental* moral importance. If we set aside cases where the promise was secured by deception, or where the promisor promised to do something immoral, or where the promisee has already released the promisor from her obligation, and so on, then for most people the fact that Jones has made a promise is a reason of intrinsic significance (a reason of fundamental importance, a reason which counts in the very last analysis) in favor of Jones’s doing what she promised. Even on this view, of course, the reason constituted by the promise may sometimes be outweighed. But it counts. *It*, the promise, counts ultimately, in a way that the existence of a student-teacher relationship does not count ultimately on the question of whether two parties should have a sexual relationship.

So, the rule that one should keep one’s promises (as it is commonly understood) is not an indicator-rule. It is not an indicator-rule just because it makes reference to a fact that figures in the ultimate moral analysis. An indicator-rule, in contrast, makes reference to facts which do not matter ultimately (like the existence of a student-teacher relationship, in the present context) but which indicate (with varying degrees of reliability) the existence of facts which do matter ultimately (such as facts about the voluntariness of a sexual relationship). Hence the name, *indicator*-rule.
When we act on an indicator-rule, our action is triggered by perception of a fact that does not matter ultimately (such as the existence of a student-teacher relationship). The perception of the non-ultimate fact triggers action because we believe it is an adequate indicator of the existence of a fact which does matter ultimately (nonvoluntariness). In a real sense we act on our belief in the existence of the ultimately important fact; but we decline to inquire further into the existence or nonexistence of that fact. The indicator-rule moves us to action because of its character as an indicator; its ruleness lies in its functioning as a bar to further inquiry.\(^{40}\)

The other feature of indicator-rules which reveals that they have no intrinsic importance is actually something that follows from what we have just been discussing. It is this. If at any time one discovers a new indicator-rule which is a better guide for dealing with some problem than one's old indicator-rule, then one ought to abandon the old indicator-rule for the new one. This change of indicator-rule need not reflect any change of moral view, nor any change in one's moral circumstances which is itself of intrinsic significance. It may just be that one has happened upon a better guide to action, and when one happens on a better guide to action, one should be guided by it.

Notice that I speak of "happening upon" an improved indicator-rule. I do this to make it clear that the agent who uses an indicator-rule is not at all times actively looking for improved rules. If one attempted to reevaluate the optimality of an indicator-rule from the ground up every time one was called upon to make a decision to which the rule was relevant, one would lose the benefit of having the rule. Therefore, one does not engage in this constant, active, reevaluation. Nonetheless, one may, in some circumstances, just run up against an obvious improvement on one's rule, which circumstances force on one's attention; and when this happens, no benefit is lost if the rule is revised. (Actually, the agent ought to subject his indicator-rules to active reevaluation from

---

40. For the benefit of the sort of reader who reads the footnotes as she goes along, it may be worth mentioning that the distinction I make here between indicator-rules and (for example) the rule that one should keep one's promises is related to Raz's distinction between reasons for belief and reasons for action, of which I shall say more presently. See infra text accompanying notes 64-79. It is also closely related to a distinction I have drawn elsewhere between "indicative reasons" and "intrinsic reasons." Regan, Law's Halo, 4 SOC. PHIL. & POL'Y 15, 20-21 (1986). To recognize something as an indicative reason is not the same thing as making that something the determining factor of an indicator-rule (nor shall I stop to detail the connections and differences), but the present discussion of indicator-rules ought to shed some light on my earlier discussion of indicative reasons.
time to time, but not on every occasion of use; and again, if the reevaluation turns up a better rule, the old one should be abandoned for the new.

What the preceding two paragraphs reveal is that indicator-rules are, in a sense, provisional. Their provisionality goes much deeper than the susceptibility to being overridden that we find in the rule that promises must be kept, and that we refer to by saying that rule holds only "prima facie." When the rule about promise keeping is overridden, it is overridden by considerations of some unrelated sort. Thus, it may be necessary to break a promise in order to help save a life. But when this is so, it does not mean, even in the precise circumstances, that the obligation to keep the promise does not count. It does count; it just does not count as much as the obligation to save life. In the case of an indicator-rule, however, the provisionality reflects the standing possibility that the indicator-rule will be found not to represent as well as some alternative indicator-rule the very considerations the indicator-rule is meant to be an indicator for. When this happens, the original indicator-rule has lost all significance. It has no further part to play in moral decisionmaking. It is superseded for all purposes by the improved rule. This is quite unlike the way in which the rule about promise keeping is "superseded" in a particular case by the obligation to save life.

In a moment I shall expand a bit on the provisional nature of indicator-rules, but first let me sum up the discussion so far. Remember that I have been trying to explain how indicator-rules fall somewhere between rules of thumb (or what I now think most people understand under the rubric "rules of thumb") and rules simpliciter (like the rule that one should keep one's promises). Indicator-rules are more substantial than rules of thumb (as many people conceive them) because an agent may be bound as a conscientious decisionmaker to adopt and act on a certain indicator-rule. Indeed, every agent is bound as a conscientious decisionmaker to adopt and act on the best indicator-rules available to her for promoting values which she is bound (independently of her desires or preferences) to promote. (Also, an agent may be bound to adopt and act on a particular indicator-rule even though it is fallible and is known to be fallible, if it is the best available.) On the other hand, indicator-rules are less substantial than rules simpliciter in that they do not focus on facts of ultimate moral significance, and they are therefore always provisional in a strong sense.

We might summarize by saying that the agent may be bound as a conscientious decisionmaker to adopt and act on a particular indicator-
rule, but that even so, the indicator-rule is not binding in itself. I intend this to sound mildly paradoxical. I hope the air of paradox will help to fix in the reader's mind a conception of indicator-rules which has always seemed straightforward and obvious to me, but which may be more idiosyncratic than I used to assume.

Now, a collage of further comments on the provisional nature of indicator-rules. First, I have said that one should abandon an old indicator-rule, even if one was bound to use it in the past, any time one finds a better one. A closely related observation is that one should ignore an indicator-rule in a particular case if one just happens to be able to see clearly that in that case it leads to a wrong result, or that the normal justification for it does not obtain. For example, a forty-year-old professor who has always followed the rule "Never have sexual relations with your students" might suddenly detect mutual attraction between himself and a student who is of the same age as himself, a mature and stable personality, well-settled in life economically and professionally, and taking the professor's course as a special student, not even pursuing a degree. In these circumstances, the reasons to fear that sexual relations might not be voluntary on the student's side essentially disappear, and the professor can quite reasonably ignore his long-followed indicator-rule.

Someone who does not like the idea of rules being disregarded might object that the old indicator-rule never really covered a case like this at all. That may be true in a sense—certainly the professor can recognize the exceptional case as exceptional without feeling that he has changed his mind about anything or has abandoned a previous moral commitment. Nonetheless, I think we misrepresent the situation if we just say the old indicator-rule did not cover this case. The case comes within the literal terms of the old indicator-rule as the professor would have stated it to himself (and quite possibly often did state it to himself). And the professor, we may assume, never made any reservation for a case such as we have imagined. He just never thought about this case until it arose. I think we must just admit that our indicator-rules are always provisional, even in the sense that we may be led to make case-by-case exceptions, and leave it at that. (There is also always something provisional in our understanding of the application to cases of genuinely bedrock moral rules, rules such as "keep your promises." But that is provisionality of a different sort.)

The more worrisome challenge to my account of indicator-rules may come from someone who admits the provisional nature of indicator-
rules, and who in fact thinks indicator-rules are so provisional that they can make no real contribution to decisionmaking at all. If we must always stand ready in principle to revise our indicator-rules or to recognize a particular exceptional case, what is left of the rule? Are we not always looking through the indicator-rule to the underlying considerations? I think the answer here is definitely, no. Even provisional rules can contribute to decisionmaking. Standing ready to abandon or ignore the indicator-rule if circumstances come to one's attention that make such treatment of it appropriate is not the same as treating the indicator-rule as perfectly transparent in every case.

We can make it clear that there is a difference here by borrowing (and slightly altering) an example from Raz. Imagine that Jones and I are engaged in a project which periodically requires the adding up of long sums by hand. Jones is a faster and more accurate calculator than I, so I have adopted the indicator-rule "Let Jones do the adding, and trust his results." Imagine now that I just happen to notice on some occasion that Jones has added up a long column of even integers and got an odd integer as the sum. I could notice this without doing the sum myself; indeed, I might be most likely to notice it if I am just abstractedly gazing at the numbers and toying with them as I wait for Jones to do the sum. But if I do happen to notice this, then I know quite definitely that in this case Jones got it wrong. So, even though I begin with no intention to do anything but follow my rule, the fact may simply force itself upon my consciousness that the rule should be ignored in some particular case. (Whether I should now do the sum myself or just ask Jones to do it over; whether I should correct Jones in this case alone or reduce my reliance on him in general; these are further questions we need not attempt to answer.)

This is a very simple-minded case. But it shows how I can discover a need to modify or ignore my current indicator-rule on some occasion without treating the indicator-rule as transparent in general, that is to say, without having the intention of looking through the rule to the underlying considerations on every occasion. More complicated cases may be much more complicated; and they would also bring it to our attention that once we admit we can disregard the rules, or revise them, in "clear" cases, then we must admit that the real issue is not absolute clarity, but sufficient clarity. Still, the more complicated cases do not differ in their essence. They do not undermine the distinction between

41. J. Raz, supra note 1, at 62.
being willing to revise or disregard the rules in certain appropriate circumstances and treating the rules as transparent in general. (And I trust it is obvious that occasional self-initiated review of one's rules for soundness is also consistent with treating the rules as non-transparent between reviews.)

In sum, indicator-rules can be always provisional without being perfectly transparent. Indeed, they can be almost perfectly opaque. But they cannot be perfectly opaque, given their nature. Because they are not significant intrinsically, they can never exclude entirely the possibility that what is significant intrinsically will rise up and manifest itself unmistakably. If it does so, it will displace the mere indicators.

One last point. Perhaps it seems that we ought to be able to pin down the degree of transparency or opacity of the indicator-rules we have been discussing by means of higher-order rules about what circumstances justify one in abandoning or ignoring various first-order indicator-rules. This is a project we can begin, but not one we can complete. I do not deny that there may be useful higher-order indicator-rules. But we cannot have a complete and inviolable system of such rules which is totally nonprovisional, totally opaque. The reason is that there cannot be a finitely-statable decision procedure which is the best decision procedure for us to follow in all contingencies. What this means is that the mode of operation of indicator-rules, as seen by the agent who uses them, must preserve a kernel of mysteriousness which cannot be eliminated. One does not ignore an indicator-rule whenever it might be wrong, nor even whenever one has some positive evidence that it is wrong. (One may have some evidence that the rule gives the wrong result in a particular case, without that evidence being strong enough to overcome one's confidence in the rule, neither by making one act against the rule, nor even by moving one to look behind it to the whole range of underlying reasons.) Just when one does ignore the rule is something we cannot specify finitely and completely. That is not, however, an objection to the model of indicator-rules. It is the price of caring about results and remaining potentially responsive to whatever situation the universe might confront us with.

Let me now summarize what we have learned about indicator-rules: (1) the best indicator-rules available to an agent may be fallible; (2) indicator-rules can be opaque to a significant degree (that is, they can make a genuine contribution of their own to the process of deciding) even though

---

they are always subject in principle to revision or to being ignored in particular cases; (3) an agent is bound as a conscientious decisionmaker to adopt (provisionally) and to act on (absent the realization of the proviso) the best indicator-rules available to her for promoting genuine values; and (4) the agent is bound to adopt and to act on these indicator-rules even though the indicator-rules, which are never more than indicators, are not binding in themselves. 43

Now back to authority. As I said much earlier, I think Raz's three theses on authority—the normal justification thesis, the dependence thesis, and the preemptive thesis—are satisfied by what I shall call the "indicator-rule conception of authority." On this conception of authority, an entity is an authority for some agent if the agent ought to adopt as an indicator-rule the prescription "Follow the directives of (the entity in question)." Alternatively, an agent regards an entity as authoritative for her if she does adopt an indicator-rule directing her to follow the directives of that entity.

Let us see how an agent might justify such an indicator-rule, specifically about governmental authority. Suppose that I am a believer in agent-neutral act-consequentialism and I say to myself:

By and large, the governmental agencies in my society are run by well-meaning people, who have the general good at heart and who have access to much better information about a wide range of problems than I do. Over a wide range of cases, when there is a law or other governmental directive that purports to control my behavior, I shall do better at doing what is morally required of me (required by reasons quite independent of the existence of the directives) if I follow government's directives than if I try to figure out what to do on my own. Therefore, I shall adopt an indicator-rule of following the directives in that range

43. I want to mention somewhere, though there has not seemed to be any perfectly apt place, the very useful discussion in Gans, Mandatory Rules and Exclusionary Reasons, 15 PHILOSOPHIA, 373, 390-94 (1986), of what Gans refers to as the distinction between "not acting for a reason" and "not considering a reason." Either Gans' distinction or a closely related distinction is crucial to my idea of how indicator-rules work. Indicator-rules are not only indicators of certain reasons, but also (indicators of) reasons for not considering certain other reasons. They are not reasons for not acting on other reasons which we know about, and can evaluate, without active inquiry. At least, they are not reasons-for-not-acting-on-other-reasons in a strong Razean sense. See Raz, supra note 38, at 221-23. They will still, of course, prevent us from acting even on other reasons that we know all about, if, in their primary indicative capacity, they are taken to indicate stronger reasons than those other reasons. So, the indicator-rule may undermine the effectiveness of a consideration which points against the rule in two different ways: (1) by preventing us from finding out about it or evaluating it, or (2) by being supposed to indicate stronger considerations in favor of the result the rule requires. This duality of function resembles the dual functioning of Raz's "reasons for not acting on other reasons." But I think there is nothing here that is exactly what Raz has in mind. He wants a stronger sort of preemptive force.
of cases where I think the government's judgment is normally better than mine.

If I adopt this indicator-rule, then I am, in an obvious sense, treating government as an authority on what I should do. Furthermore, the relationship between government and me satisfies Raz's three theses about authority.

First, my ground for treating governmental directives as authoritative (for adopting my indicator-rule) is essentially the same as Raz's ground for saying government is an authority under the normal justification thesis. It may not be precisely the same. Raz says an entity is an authority for an agent under the normal justification thesis if the agent would be well-advised to treat the directives of that entity as "authoritatively binding." Raz may mean by "authoritatively binding" something more than the "bindingness" of indicator-rules. If he does, then the test for authoritativeness under Raz's normal justification thesis is slightly different from the test I used in deciding to adopt an indicator-rule of following the authority. I shall say a bit more about whether Raz's test and mine differ in this respect when I discuss the preemptive thesis. For now, let two observations suffice. First, if we are trying to characterize authority in a useful and revealing way, then the question of what to mean by "authoritatively binding" is part of the problem, not one of the givens with which we must operate. If, with that point in mind, we take Raz's phrase "authoritatively binding" to mean just "binding in whatever way it is appropriate to regard authority as binding," then we have as yet no reason at all to think Raz's test differs from mine. Second, even if there is some difference here, the structural similarity between my suggested argument for adopting the indicator-rule and Raz's normal justification thesis is much more impressive than any difference.

Next, my agent-neutral consequentialism entails that anyone who issues directives purporting to control my behavior ought to base those directives on reasons that are independently applicable to me, simply because the same basic reasons apply to both of us. In other words, my argument for authority entails the dependence thesis. So far as I can see, this argument from agent-neutrality is also Raz's reason for accepting the dependence thesis. Raz can hardly be said to argue for the dependence thesis—most of his argument about it takes the form of answering possible objections. Nor does Raz ever take a definite position on agent-neutrality. But the flavor both of the book as a whole and of the section

44. J. Raz, supra note 1, at 53.
entitled "Separateness of Persons: Agent-Neutrality" seems more neutralist than otherwise. Most importantly, neutralism seems the most plausible foundation for the dependence thesis: If we are not agent-neutralist, then, absent some special finagling of our theory, someone who purports to direct my behavior ought to be guided, not by reasons which are independently applicable to me, but by whatever reasons are applicable to him. After all, his activity of directing my behavior is just as much his activity as anything else he does, so it should be guided by reasons-for-him, which may be different from reasons-for-me.

Finally, if I adopt the indicator-rule "Follow governmental directives," then knowledge of a relevant directive will (normally) preempt me from acting on my independent judgment about what to do. I am not preempted because I am bound in any deontological sense. (Remember my claim that I am bound to adopt the indicator-rule, but it is not binding in itself.) Rather, I am preempted because the whole point of having an indicator-rule is to rely on it instead of relying on direct calculation. The professor who adopts an indicator-rule against sleeping with his students has decided precisely not to consider case-by-case whether some student (some "standard" student, we might say, remembering the exceptional case of the mature and independent student) might enter a sexual relationship voluntarily. Similarly, if I decide to leave the arithmetic to Jones, then I have decided precisely not to do it myself. So, what Raz's preemption thesis says about authoritative utterances, we can say about any indicator-rule (including an indicator-rule directing me to follow a putative authority): "[T]he fact that [an indicator-rule] requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them."47 Note in particular that, just as Raz is at pains to point out that an authoritative directive does not lose its authoritativeness just because it happens to be mistaken,48 so we have observed that even the best available indicator-rule for dealing with some problem may lead us into a certain number of mistakes in particular cases.

Returning now to a point I raised earlier: Just as Raz might mean, or want to mean, something more by "authoritatively binding" than the bindingness of indicator-rules as I understand them, so Raz may mean,
or want to mean, something more when he says authoritative utterances "exclude" other reasons than I mean when I say the same thing about indicator-rules. I have insisted that indicator-rules cannot be perfectly opaque to the reasons on which they are based. Raz may think that authoritative utterances are perfectly opaque. Of course Raz admits that one may ignore an authoritative utterance in certain circumstances. One may ignore it, not merely when it is ultra vires, but also, for example, if one knows the person in authority was drunk when she issued the directive, or if one has new information which one knows was not available to the authority when it issued its directive. But such considerations all go to show that the authority's deliberations which led to the promulgation of the directive were defective. Raz might still want to claim that, unless there is direct evidence of the defectiveness of the authority's deliberations, the subject of the authority is absolutely preempted from deciding for himself on the basis of the underlying reasons.

I say "direct" evidence of the defectiveness of the authority's deliberations because, of course, any time one suspects the authority's directive leads to a wrong result in a particular case, one has some indirect evidence for the defectiveness of the deliberations, in some respect and in some sense. (And I have just said "in some respect and in some sense" because if we looked closely at the notion of "defective deliberation" we would find significant choices to be made about what we mean by that.) Still, despite all of this uncertainty, the claim that the subject of authority is absolutely preempted from looking at the underlying reasons for an authoritative directive unless there is direct evidence for the defectiveness of the authority's deliberations sounds like something Raz might want to say about authority, and which I would not say about indicator-rules. Neither the case of ignoring Jones's arithmetic when he adds up a string of even numbers and gets an odd number, nor the case of the professor having sex with his mature and independent student can be thought of as involving direct evidence of defective deliberation in any ordinary sense.

49. Id. at 42.

50. It is appropriate to make here a point about authoritative directives that I made earlier in the text about indicator-rules. The issue we are concerned with is transparency, not ability to override. Thus, Raz recognizes explicitly that an authoritative directive might be intended to summarize only some of the reasons relevant to a particular action, not all of them; and if the authoritative directive is so intended, then of course it can be overridden (in an obvious sense) by consideration of reasons it does not purport to represent. Id. at 46. In this respect, Raz's view of authoritative directives and my view of indicator-rules are identical. The issue on which we may disagree is the circumstances under which one may look through a directive or an indicator-rule to the very reasons it is intended to represent.
(In the case of Jones, we have absolutely conclusive evidence that something was wrong with his deliberation, but the evidence is not direct. In the case of the professor, if we say the deliberation which led him to adopt his indicator-rule against sex with students was defective because he failed to notice the possible case of the mature and independent student, then we can be quite certain in advance that any instance of deliberation by anyone about a problem of any significant complexity is defective in some respect, and the criterion ceases to be restrictive.)

I say Raz might want to make the strong claim I have described about the exclusionary force of authoritative utterances. I do not think he should want to make it, and it is unclear whether he really does. At one point Raz is explicitly undecided. Remember that the example of Jones's egregious arithmetical error was borrowed from Raz. In the paragraph from which the example was taken, Raz's main object is to distinguish between the notion of a clear mistake and the notion of a mistake with significant consequences. But Raz raises in passing the question whether we are bound to follow authoritative utterances which are clearly in error (as Jones is when he gets the odd sum). And he explicitly declines to express an opinion.  

Even if Raz is ambivalent, that raises a question—Why is he ambivalent on a matter that seems to me so clear? One possible explanation we shall encounter later on, when we recur yet again to this arithmetical example of Raz's in connection with the question of whether one owes authority "obedience." Another possible explanation is that Raz wants authoritative directives to mediate a sort of Rawlsian "overlapping consensus," and he may think they can accomplish this only if they are regarded as "narrowly absolutely exclusionary" (that is to say, absolutely exclusionary unless they are ultra vires or the result of defective deliberation). This raises very complicated issues which I cannot discuss here about the functioning of rules in societies whose members have fundamentally different moral views. I have discussed this matter briefly elsewhere (not as fully as it deserves). For the present, I can only say that, although the need to "share" rules with people of quite different moral views may give one new reasons to follow the rules, it does not justify

51. Id. at 62.
52. See infra text accompanying notes 77-79.
53. J. RAZ, supra note 1, at 58.
54. See D. REGAN, supra note 42, at ch. 4 (note especially the comments in the second paragraph of the chapter on the range of problems for which the punishment problem is a stand-in).
treating the rules as different in their fundamental operation from indicator-rules as I have described them. It does not justify treating the rules as "narrowly absolutely exclusionary."

There is yet another possible explanation for Raz's ambivalence which deserves brief discussion. When we think about the relationship between authority and subject, we can think about it from the subject's point of view or from the authority's point of view. My remarks in this essay are relentlessly from the subject's point of view. I think that from the subject's point of view, the indicator-rule conception of authority (modified in a way that we shall see is necessary to account for the role of authority in solving coordination problems) is the right conception. On the other hand, there is much reason to think that if we are trying to capture the authority's point of view, the indicator-rule conception is not quite so apt. To put it crudely, the subject, as a moral agent, cannot give up in principle the right to make exceptions or alterations to the authority's directives when and as they are seen to be required by the underlying reasons that justify deference to the authority. The authority, on the other hand, if it does not fully trust the subjects, would like to deny them any recourse to the underlying reasons at all.

It seems to me that Raz's predominant focus in *The Morality of Freedom* is on authority from the subject's point of view. That is why I focus entirely on the subject's point of view. But Raz does occasionally consider the authority's point of view, and this could account for some ambivalence (if I am right in detecting some) about the degree of preemptiveness of authority.

The point that authority may look different from the subject's point of view and from the authority's point of view seems to me extremely important, although I shall make no more of it in this essay. In the end, of course, we want to be able to unify the two points of view somehow. One reason is that as citizens of a democracy we must be able to take both points of view. But before we try to unify the viewpoints, we ought to understand each properly in itself. Hence, I shall continue to concern myself solely with the subject's point of view.

55. See, e.g., J. Raz, *supra* note 1, at 76-77

56. I owe the point partly to Fred Schauer, and partly to joint discovery in the course of highly illuminating discussions with Fred Schauer.
C. Raz’s Objections (Herein of Obedience and Coordination Problems)

I have now explained the indicator-rule conception of authority, and I have shown that this conception of authority satisfies Raz’s three theses, or at least, reasonable interpretations of those theses. Yet Raz rejects the indicator-rule conception. Why?  

First, let us notice a feature of the indicator-rule conception which distinguishes that conception strongly from a “standard” or “commonsense” conception of authority. We have discussed, for purposes of exposition, the indicator-rule “Follow authority.” But it is very likely that for some individuals, the optimal rule is not “Follow authority,” but rather “Follow authority with respect to (some particular subset of the problems authority addresses).” If I am better informed than the authority on some particular class of issues, then quite possibly I should not extend my indicator-rule about following authority to that class of cases. It turns out that the authority is authoritative for me with respect to only some of its directives. A further consequence of the same idea is that the authority may be authoritative over different ranges of issues for different people. (Notice the point here is not just that different people might actually recognize the authority in different ranges of cases, but that the optimal indicator-rules for different people might require them to follow authority in different ranges of cases.) In sum, the authority is authoritative in differing degrees, and with respect to different issues, for its various subjects.

This conclusion is totally at odds with the standard, common sense conception of political authority. The indicator-rule conception of political authority is highly nonstandard. It is therefore worth emphasizing that this is not any part of Raz’s objection to the indicator-rule conception of authority. Raz recognizes the varying authoritative status of authority as a consequence of his own view of authority, and he accepts it.  

(Incidentally, this is one of the reasons I see Raz’s discussion as concerned predominantly with the subject’s viewpoint. The authority’s viewpoint is unlikely to admit this varying authoritative status.) There is an obvious common sense objection to the claim that government may

57. What Raz explicitly rejects is a conception of authority he refers to as the “recognitional” conception. J. RAZ, supra note 1, at 28-31. For a number of reasons, it is unclear whether Raz’s recognitional conception is identical to my indicator-rule conception. But it is clear that Raz would regard his principal reasons for rejecting the recognitional conception as applying to the indicator-rule conception as well.

58. Id. at 74, 77-78.
not be authoritative for all its citizens over the entire range of problems it addresses: this claim seems to put in jeopardy the right of even a just government to use coercion in support of all its commands. But Raz, quite rightly, distinguishes the issue of authoritativeness from the issue of the justification of coercion. Government may justifiably coerce in some cases where it lacks authority.\footnote{59. Id. at 103-04. Cf. A. J. SIMMONS, MORAL PRINCIPLES AND POLITICAL OBLIGATIONS 195-201 (1979).}

So, Raz cannot object to the indicator-rule conception of authority on the ground that it makes government authoritative in different ways for different people. That is part of his own view. Raz’s principal actual objections to the indicator-rule conception of authority are two: (1) The indicator-rule conception of authority does not justify our thinking of the relationship of the subject to the authority as one of obedience; it does not give rise to an “obligation” to “obey.”\footnote{60. J. RAZ, supra note 1, at 29, 60.} (2) The indicator-rule conception of authority does not allow us to account for government’s contribution to solving coordination problems.\footnote{61. Id. at 31, 49-50.}

Now, the first objection seems to me true, but not an objection at all. I agree that the indicator-rule conception does not support talk of “obedience,” nor of “obligation” in any sense which is different from mere “ought.” (I talked at one point of subjective and objective obligation in connection with indicator-rules, and I talked of indicator-rules as being “binding” in certain senses. But these uses of “obligation” and “binding” can be reduced to mere talk of “ought.” Raz wants authority to impose obligations in some stronger sense.) We may follow an indicator-rule or not; indeed, we may be morally required to follow it, or not to; but even when we follow it, we do not obey it. Similarly, even if we have the indicator-rule “Follow authority,” our adopting and acting on that rule leads us to be guided by the directives of the authority, but not to obey them.\footnote{62. See infra text accompanying notes 208-15 (especially the final paragraphs of the Appendix).} But why is this an objection to the indicator-rule conception? Why should we not dispense with the language of “obedience” and “obligation” if that is the direction in which Raz’s three theses lead? Perhaps there are reasons to think in terms of obedience which we have not yet considered. But I see no reason to make it a fundamental criterion for the correct account of the relationship between government and citizen that it include the notion of obedience. If we do not, then Raz’s first objection is not, by itself, an objection at all. (If it be objected that
there is a criterial connection between "practical authority" and obedience, I would respond that we should not make it criterial for an analysis of the relationship between government and citizen that it show the relationship to involve "practical authority" so understood.

Raz's second objection is also true, and it is a good objection. For reasons I shall explain presently, the indicator-rule conception of authority does not allow us to account for government's role in solving coordination problems, and it is a criterion for an adequate account of authority that it should allow us to account for such an important governmental function. However, Raz makes too much of this second objection. Specifically, Raz assumes that there is no way to revise the indicator-rule conception of authority so that it can account for government's role in solving coordination problems without turning it into a conception of authority which involves notions of obedience and the like. About this, Raz is mistaken. The indicator-rule conception of authority must be supplemented if we are to account for government's role in solving coordination problems. But the required supplementation will neither bring in the notion of obedience nor suggest that the indicator-rule account is inadequate where coordination is not in issue.

The connection Raz sees between coordination problems and obedience is mediated through the concept of a "reason for action," and specifically through claims about the status of authoritative utterances as reasons for action. So, before I discuss government's role in solving coordination problems, which turns out to be the crucial issue between Raz and me, I must pause over reasons for action.

Raz distinguishes between reasons for belief and reasons for action. (Indeed, we might have listed as a further Razean objection to the indicator-rule conception of authority, that it turns authoritative utterances into mere reasons for belief instead of reasons for action. But in the end I think this objection is subordinate to the two we listed.) Now, I agree that there is a significant distinction here, but we must be careful about just what it is. There is both a strict sense and a loose sense of "reason for action," and even reasons for belief can be reasons for action in the loose sense. There is nothing the least odd in my saying, "I'm going upstairs to watch television because Vince told me there's a Tigers game on." If I say this, I am justifying my action, not merely explaining it; and I am offering Vince's assertion as a reason for my action, which in some sense it certainly is. Strictly speaking, however,

63. J. Raz, supra note 1, at 28-31.
64. Id. at 29.
Vince's assertion merely gives me reason to believe there is a Tigers game on. The reason for action, in the strict sense, is that there is a Tigers game on, or, less elliptically, that "If I go upstairs, I will get to see a Tigers game." Reasons for action in the strict sense are facts which count, in themselves, by virtue of bedrock moral truths, in favor of or against actions. That Vince says there is a Tigers game on; even that I think there is a Tigers game on; these are not facts that count in themselves. What counts in itself (because watching a Tigers game would be valuable) is that there is a Tigers game on and that if I go upstairs I will get to see it. Vince's assertion is a mere indicator for this fundamental fact. An indicator is a reason for action in the loose sense, but only the fact it indicates, the fact that counts in itself, is a reason for action in the strict sense.\(^\text{65}\)

Now, what about authoritative utterances? Raz claims authoritative utterances must be reasons for action.\(^\text{66}\) Plainly they must be reasons for action at least in the loose sense. But about that there is no problem. The indicator-rule conception of authority treats authoritative utterances primarily as reasons for belief (reasons for belief, we should note, that have, in addition, special operational consequences for the decisionmaking process because they are relied upon to limit further active inquiry into other possible reasons for or against the relevant belief, or for or against action); but even on the indicator-rule conception, authoritative utterances are reasons for action in the loose sense, as we have seen reasons for belief often are.

In fact, Raz means to claim that authoritative utterances must be reasons for action in the strict sense. If this were true, the indicator-rule conception of authority would be refuted, because on the indicator-rule conception of authority, authoritative utterances are not reasons for action in the strict sense. But I think the premise is false. There is no ground for thinking authoritative utterances must be reasons for action in the strict sense. One reason Raz suggests for thinking authoritative utterances must be reasons for action in the strict sense is that unless they are reasons for action in the strict sense, our relation to authority is not one of obedience.\(^\text{67}\) But I have already said that I see no reason to assume obedience must come into the picture. So this is no real argument. Raz's only substantial argument that authoritative utterances

---

\(^{65}\) What I am now calling a "reason for action in the strict sense" is what I called an "intrinsic reason" in Regan, supra note 40, at 20-21. "Reasons to believe," when the beliefs are relevant to practical questions, are what I there called "indicative reasons."

\(^{66}\) J. RAz, supra note 1, at 29, 31.

\(^{67}\) Id. at 29.
must be reasons for action in the strict sense is that, unless they are, we cannot account for government's contribution to solving coordination problems. 68

So here we are, back at the crucial issue. In what follows, when I refer to "reasons for action" without further qualification, I shall mean reasons for action in the strict sense.

Lest the reader get lost right at the outset, let me give a quick precis of the argument ahead. Specifically, let me remind the reader that there are really two issues before us, which Raz runs together. One issue is whether the indicator-rule conception of authority can account for government's role in solving coordination problems. The other issue is whether, in order to solve coordination problems, authoritative utterances must be reasons for action (in the strict sense, remember). Raz thinks that in order to solve coordination problems, authoritative utterances must be reasons for action; that on the indicator-rule conception authoritative utterances are not reasons for action (this is true); and that that is why the indicator-rule conception cannot account for government's role in solving coordination problems. I agree with Raz that the indicator-rule conception as we have developed it so far cannot account for government's role in solving coordination problems. But the reason is not that in order to solve coordination problems, authoritative utterances must be reasons for action. Rather, what must be true of authoritative utterances if they are to solve coordination problems, is that they must be able to change our reasons for action (still in the strict sense). The basic objection to the indicator-rule conception is that it does not even allow us to account for how authoritative utterances do this. But it is not necessary that authoritative utterances be themselves reasons for action, if they can change our reasons for action without being reasons for action. I shall argue that they can.

Let me emphasize something about my last claim by putting the point in Raz's terminology. I am not asserting the "no difference thesis." I agree with Raz that the exercise of authority sometimes makes a difference to what its subjects ought to do, and I agree with Raz that it is an objection to the indicator-rule conception of authority that it cannot account for this in the appropriate sense. What I am asserting is that we can produce an account of authority, a supplemented version of the indicator-rule conception, that does account for how the exercise of authority can make a difference to what its subjects should do, without bringing

68. Id. at 31.
in the idea that authoritative utterances are themselves reasons for action in the strict sense. (And because the supplemented conception of authority will not treat authoritative utterances as reasons for action, it also will not support notions like obedience.)

And now, finally, to the real discussion. A central function of authority in general, and of governmental authority in particular, is the solution of coordination problems. Raz's main objection to the indicator-rule conception of authority turns out to be that it cannot account for how authoritative utterances affect our reasoning about coordination problems. The argument, which I concur in up to a point, is this: Imagine that there is no law (and no existing convention) about whether to drive on the right side of the road or the left. In this situation, all drivers can recognize that it would be good if everyone drove on one side or the other, but even so, no driver has any reason (in general) to pick one side over the other. Our reasons for wanting to have a general practice simply do not specify the content of the practice with sufficient definiteness to bring it into being. Now, suppose government speaks: "Drive on the right." From this moment, everyone has a reason to drive on the right, and coordination will be achieved. Before government spoke, I had no reason to drive on the right as opposed to the left. After government speaks, I do have such a reason. Obviously, government's speaking has changed my reasons for action. But indicator-rules merely summarize reasons for action; they do not create or destroy them. Therefore, the authoritative utterance "Drive on the right" has done something that the indicator-rule conception of authority cannot account for. (Let me expand just a bit, for the reader who is still doubtful. A standard instance for the indicator-rule conception of authority might be a government directive banning the use of some dangerous chemical. Such a directive need not be thought of as adding to the reasons for not using the chemical that existed independently of the directive. The directive may inform us of the existence of the reasons for not using the chemical, but it does not change them. In contrast, the directive "Drive on the right" does add to the pre-existing reasons for driving on the right. Before government spoke, there was no reason to prefer the right to the left. Now there is.)

So far Raz and I are in agreement. We agree that the authoritative utterance "Drive on the right" changes my reasons for action, and we agree that the indicator-rule conception of authority cannot account for this. It is when we go on to ask how we do account for this effect of authoritative utterances that Raz and I part company.
Raz thinks the authoritative utterance changes my reasons by being itself a reason. Before government spoke, I had no reason to drive on the right. After government speaks, I have such a reason. What more obvious than to suppose that my new reason just is that government has spoken, or in other words, that the authoritative utterance is itself a reason to act?

This may be the obvious answer to our question about how authority creates a new reason in the coordination case, but I do not think it is the right answer. I am perfectly willing to admit, of course, that the authoritative utterance is a reason for action in some loose sense such as I have already adverted to when I said even a reason for belief could be a reason for action. But Raz is interested in reasons for action in the strict sense. What I deny is that the authoritative utterance is a reason for action in the strict sense, even in the coordination context.

The correct view, it seems to me, is that the authoritative utterance changes my reasons for action (in the strict sense) without itself being a reason for action (in the strict sense). This is a possibility Raz just seems to overlook. How does the utterance accomplish this? We can describe the process in either of two ways.

If we look at things just from my own, individual point of view, we can say the following: Before government speaks, other drivers will choose which side to drive on at random (so far as I can tell), so I have no reason, given my goal of avoiding accidents, to prefer one side or the other. After government speaks, others will tend to drive on the right, so I will minimize accidents by driving on the right, and that is my reason (in the strict sense) for doing so. On this account, the authoritative utterance gives me a new reason for driving on the right (namely, that doing so will avoid accidents), by altering other people’s behavior. The authoritative utterance causes a change of circumstances, and the new circumstances include a reason for driving on the right that the old circumstances did not. But notice also that on this account the contribution of the authoritative utterance to the change in my reasons depends on its causal efficacy. If the authoritative utterance had no effect on others' behavior, then it would have no effect on my reasons for action. It follows that the authoritative utterance is not a reason in itself.

There are a number of closely related grounds for discomfort about this way of looking at things. Let me mention the subtlest first. Raz might say that only legitimate authoritative utterances are reasons for
action and that in coordination contexts recognition as an authority is a normative prerequisite to legitimacy. But (Raz might continue) if the authority which says “Drive on the right” is recognized, then its directive will necessarily be causally effective (since that is part of what recognition means). So a legitimate directive to drive on the right will be effective, and I cannot argue that such a legitimate authoritative utterance is not itself the reason for action by pointing to the possibility that the authority might speak without anyone taking heed.

I have no objection to Raz’s desire to connect recognition and legitimacy (if “legitimacy” matters at all) nor to his claim that only legitimate authoritative utterances are reasons for action (if any are). Even so, I think the argument I have just imagined would be captious as an answer to the claim that the real reason for action is not the authoritative utterance but rather the consequent change in others’ behavior. After all, the point of the imagined argument is just that if a legitimate authority speaks, behavior will change. It says nothing against the claim that the change in others’ behavior provides the real reason for action. Still, perhaps what I really want to say is not that the imagined argument would be captious, but that it would be captious unless it were intended to raise a more substantial line of objection, to which I now turn.

Someone might object that my account of how the governmental directive changes my reasons for action is either incomplete or circular. It is all very well to say that my reasons for action change in the face of others’ changed behavior. But why do those others change their behavior? Until we have explained the change in their behavior, we have not really explained how the government’s directive changes my reasons for action. Of course, it could just be that everybody else in society responds to government’s orders like Pavlov’s dog hearing the dinner bell. But that is not an explanation we can be satisfied with in the present context. I and the others are equally moral agents and equally subject to this authority for present purposes. We want an explanation in which I and others figure symmetrically. The obvious way to restore symmetry is to explain each other person’s changed behavior as a reason-based response to the changed behavior (including mine) that confronts him, but then the whole “explanation” seems circular.

To fill in the gap without asymmetry or circularity, we obviously need an explanation of how the governmental directive can change everyone’s reasons for action “all at once.” It may seem that at this stage the

69. *Id.* at 46.
70. *Id.* at 56.
only possible explanation is the one I am trying to avoid: the directive changes everyone's reasons at once because it is a new reason that applies to everyone involved in the coordination problem. This is in fact Raz's claim—that there is no other possible explanation. But I think there is another explanation, and a better one.

In a nutshell, governmental directives solve coordination problems by making particular solutions salient. If a number of people are involved in a coordination problem and if they understand the nature of such problems, then they know that the problem can be solved, indeed is solved, if there is anything that makes one solution "stick out." If one solution sticks out, and if everyone sees that one solution sticks out, then everyone will act on that solution, and coordination is achieved. (What it means for one solution to stick out is just that everyone sees that it sticks out, and everyone sees that everyone will see that it sticks out, and so on.) The solution that sticks out need not stick out by being better in its results than other solutions. Indeed, hard coordination problems are defined precisely by the absence of any solution that sticks out by having the best results. The classic discussion of the variety of possible modes of sticking out is in Thomas Schelling's The Strategy of Conflict. But it is not possible to catalogue exhaustively modes of sticking out, nor to specify in advance what will make particular solutions stick out for particular groups of people. Whatever works, works.

The notion of "sticking out," or salience, is more than a little mysterious. Witness the circular "definition" I gave a moment ago ("what it means for one solution to stick out is just that everyone sees that it sticks out, and . . .") and similarly my statement that "whatever works, works." But despite its mysteriousness, salience is an everyday psychological phenomenon that we have all relied on, consciously or not, to solve everyday coordination problems. When friends get temporarily separated on an outing, and join up again by figuring out the "obvious" place to meet, they have solved a coordination problem by recognizing the salience of one solution.

So, I suggest that governmental directives solve coordination problems by making particular solutions salient. If this is true, then there is no reason to regard the directive itself as a reason for action in the strict sense. The only reason for action in the strict sense is that the particular solution picked out by government now is salient. It was the governmental directive that made it salient, but even so, its being salient

71. Id. at 31.
is a different fact than its having been ordered by government. The governmental directive has caused the salience, but does not constitute the salience. (The same solution could have become salient in some other way.) Therefore, it is not the governmental directive that constitutes the reason for action.\(^7\)

Notice that in order for a governmental directive to create salience, it is not necessary even for anyone to believe that the directive is itself a reason for action. People have to be alive to possibilities for salience, which means also that they have to be aware of the practical significance of salience; and they have to have some shared psychology in virtue of which governmental directives can create salience. But that is all. The governmental directive need not have, or be thought to have, any more intrinsic normative significance than the information booths in railroad stations or in large department stores (plausible solutions to other, every-day, coordination problems).\(^7\)

---

73. For the sake of precision about my own view, let me insert an observation that is not really relevant to my argument against Raz. I say in the text that the reason for action in the strict sense is that a particular solution is now salient. I am not certain that I really want to go even that far. I might want to claim that the only reason for action in the strict sense is just that, in the face of others' salience-induced behavior, my driving on the right will avoid accidents. The general idea of a "reason for action in the strict sense" is that only facts that count in themselves for or against actions, by virtue of bedrock moral facts, are such reasons. But it is unclear without much further thought, whether salience counts, in itself. The issue involves deep questions about whether the ultimate addressees of moral commands can in some sense be collections of individuals acting in concert. For the present, then, I am not absolutely certain whether salience is a reason for action in the strict sense. But I am certain that even if salience is, the directive that creates the salience is not. That is the main point for present purposes.

74. I have admitted, what cannot in any event be denied, that there is an element of mystery in the psychological phenomenon of salience, and perhaps also in our ability to make practical use of the phenomenon. Lest Raz try to make this mysteriousness an objection to my view, I would point out that his view contains a structurally identical mystery. He says that in coordination contexts, the legitimacy of an authority may depend on whether it is recognized. J. Raz, supra note 1, at 56. It might seem that I can decide whether an authority is legitimate, and therefore whether I should recognize it, by considering whether others already recognize it. But what if others are as aware of the problem and as moral as I? They see that authoritative for them depends on recognition by others, including me. If we all see that we are in this position, we may all be able to recognize the authority together, but there is some ineliminable mystery in just how we do so. This mystery is the equivalent of the mystery of exploiting salience, if it is not the very same mystery. (Notice Raz cannot say his mystery is solved by the fact that some de facto authority has the actual power to coerce and the intention to use that power in support of its directives. Of course such power is likely to be unique, and it is therefore a factor contributing to the salience of the authority's solutions. But the mere fact that some putative authority presently has coercive power does not entail that general recognition of that authority is the only solution, or the best solution, to the problem shared by citizens of what authority to recognize. Furthermore, if the circumstances are such that brute power is thought to settle the matter, then it is doubtful we should speak of people as "recognizing" the authority at all, as opposed to merely knuckling under.)
By now the reader may think I am making a mountain out of a molehill. Does it really matter whether we believe as I do that government directives solve coordination problems by creating salience or believe as Raz does that government directives solve coordination problems by being themselves reasons for action?

I think it matters a great deal. It may or may not make much difference in practice (more of that in a moment), but it makes a great difference in the flavor of our theory, in the general color of our thinking.

If we think of governmental directives as solving coordination problems by creating salience, then we will not be inclined to think of citizens who rely on a directive to extricate themselves from such a problem as "obeying" the directive or the government. Just as was true on the indicator-rule conception of authority, the whole family of notions connected with obedience seems out of place. Remember that we got onto the discussion of coordination problems because I conceded to Raz that they represented a kind of case to which the indicator-rule conception of authority seemed inadequate. But I intimated that Raz's reasons for rejecting the indicator-rule conception (as a complete account of authority) did not justify his talk of authority as a reason for action, which he specifically links with authority as requiring obedience. In my view, if we combine the indicator-rule conception (for some functions of authority) and the salience creating conception (for other functions of authority), and if we remember that constituted powers may sometimes be justified by bureaucratic necessity in enforcing the details of rules that, in respect of their details, do not provide citizens with any reason at all for action (except by indirect reasoning about the cost and consequences of enforcement), then we have a complete account of authority (or perhaps better, of the usefulness of "authorities") that has no connection at all with ideas of obedience and the like. Subjects of authority whose view of authority is the view I have been developing will often do what authority directs. And they will do what authority directs partly as a consequence of authority's direction. But they will not regard themselves as obeying the authority. They will not regard an authoritative utterance as itself a reason for action in the strict sense; and as Raz says, unless we regard authoritative utterances as reasons for action in the strict sense, then our relation to the authority is not one of obedience.

Now, as to whether the difference between Raz and me matters in practice. This is always a hard question to answer, given the wide gap

75. Id. at 29.
76. Id.
between theorizing and deciding particular cases. Almost any plausible
view can be so applied as to justify almost any plausible conclusion in
almost any realistic case. I am tempted to say that such a large difference
of theoretical attitude as whether we connect authority with obedience
must make some difference in practice, and to leave it at that. But there
is one class of cases which may repay brief discussion.

At one point, as I have already mentioned, Raz considers in passing
whether we have a duty to obey authoritative directives which are clearly
in error (clearly in error, that is, by reference to the underlying reasons
for action which, according to the dependence thesis, should guide the
authority). Raz’s main object is to distinguish between clear error and
significant error, and to argue that authoritative directives may be bind-
ing even when they are significantly in error, provided they are not
clearly in error. I agree with Raz about the distinction and its relevance.
(Even a good indicator-rule, properly used, may on occasion lead us sig-
nificantly astray; but if it is a good indicator-rule that does this while it is
being properly used, there is no cure which would improve things on
balance.) But with regard to cases where the authority is clearly
in error (whether significantly or not) Raz “express[es] no opinion” on whether
the authority’s directives are still binding.\footnote{77}

My own view, of course, is that if the authority is clearly in error, its
directives are not binding. We do not follow an indicator-rule in a case
where it is clear, without going through the calculation or the close
inspection of our own motives that the indicator-rule is designed to obvi-
ate, that the indicator-rule leads to the wrong conclusion. For the same
reasons, we should not follow a clearly erroneous directive from an
authority.\footnote{78} On the other hand, it seems that Raz ought to have no diffi-
culty concluding that we are bound to obey an authority even when the
authority is clearly in error, at least if the error is not significant. Raz
speaks of an obligation to obey, and he purports to regard that obligation
to obey as itself a reason for action in the strict sense. How could such a
reason for action, which must have some weight of its own, not require us
to obey even in cases of clear error, provided the error is small?

It seems to me that Raz’s indecision about cases of clear error
reveals the tension in his whole analysis of authority. Raz’s three theses

\footnote{77} Id. at 62.

\footnote{78} If the authority we are talking about is a widely-recognized public authority, there may be
power-of-example considerations that are relevant to the question whether to follow a clearly errone-
ous directive of authority and that are not relevant to the question whether to follow an indicator-
rule when it is clearly wrong. But Raz neither relies on nor discusses power-of-example considera-
tions, so I shall ignore them also.
about the structure of authority point to an indicator-rule conception of authority (which requires some supplementation to deal with coordination cases), and the indicator-rule conception entails that clearly erroneous directives can be ignored. But Raz also wants to find an obligation to obey authority, and if such an obligation exists, in any substantial sense, clear but minor errors do not justify disobedience. The indicator-rule conception and the obedience conception simply do not fit together, no matter how much Raz wants them to. 79

79. For further discussion, focusing on a particular passage in Raz, of how consistently one should follow authority and whether, when one follows authority, one is “obeying” it, see the Appendix entitled “Is There a Theorem Calling for Exceptionless Deference to Authority? (And If So, Is That Obedience?)” (On the issue of whether following is obeying, see especially the final few paragraphs of the Appendix.)

In discussing whether and how we must go beyond the indicator-rule conception of authority to account for the various functions of government, I have discussed only coordination problems. Raz has some other cases that he thinks illustrate the need to go beyond the indicator-rule conception. He may think that those cases establish a need to regard authoritative utterances as reasons in the strict sense even if coordination cases do not. See J. RAZ, supra note 1, at 48-52. (One reason I put this considerable discussion into a footnote is that it is unclear how far Raz intends these other cases to support the claim that authoritative utterances are reasons for action. In the passage I am about to discuss, Raz’s main objects are to show that government can do various things consistently with the dependence thesis—as I agree it can—and to show that we should reject the “no difference thesis”—as I agree we should. The problem is that in an earlier passage that I have referred to, Raz claims that we can explain how government does certain things only by assuming that authoritative utterances are themselves reasons for action. Id. at 31. That earlier passage contains a cross-reference that refers to the passage I am about to discuss as its continuation.)

The first of Raz’s further cases is where government issues a directive that eliminates some pre-existing indeterminacy in a particular individual’s obligation. For example, perhaps it would be equally good for me to pay some tax in either quarterly or monthly installments, but authority directs me to pay it monthly because “there are decisive reasons against [leaving this choice open for every individual].” Id. at 49. In such a case, Raz argues, the authoritative directive gives me a new reason to pay monthly. Now, the main thing I want to say about this example is this: If the reason for authority’s decision is the undesirability of leaving every individual a free choice, that suggests that the underlying problem of when to pay one’s taxes is a coordination problem. (As is, incidentally, the question of how much tax to pay—but we need not pursue that.) If the underlying problem is a coordination problem, then what I have said in the text already explains how the authoritative utterance may solve the problem without being a reason for action in the strict sense. In fact, I think most, perhaps all, questions of bureaucratic convenience reflect underlying coordination problems, either between the individuals governed by the bureaucracy, or between some such individuals and the members of the bureaucracy. (There also may be coordination problems strictly between members of the bureaucracy, but these will not lead to regulations for people outside the bureaucracy.)

Raz might object that the notion of a coordination problem between individuals governed by a bureaucracy and the members of the bureaucracy is much too open-ended to be a useful category. Without conceding that claim, let me approach what I think is the underlying issue here in a slightly different way, suggested by a later episode of Raz’s discussion. Id. at 51-52. Raz points out that authoritative directives are necessarily blunt instruments. Bureaucracies may have to adopt de minimis rules; they cannot write all the exceptions into laws that might be justified by case-by-case inquiry. “The intrusion of the bureaucratic considerations is likely to lead to solutions which differ in many cases from those an individual should have adopted if left to himself.” Id. at 52. Raz states that authority may nonetheless be justified in adopting and enforcing such blunt regulations. And of
D. CONSENT TO AUTHORITY AND RESPECT FOR LAW

Up to this point, we have been treating political authority as an instance of authority in general. Raz’s three main theses about authority—the dependence thesis, the normal justification thesis, and the pre-emption thesis—all apply to all types of practical authority. And Raz’s course, I agree. But that is not the issue I am now concerned with. The issue is whether these regulations constitute new reasons for individuals to do what the regulations require. Raz is not in a position to insist in general even that they create new reasons (much less that they constitute them), since he does not believe that all laws are authoritative for all subjects, and he himself distinguishes between the question of the authority’s right to coerce and the subject’s duty to obey. Id. at 103-04. At the very least, Raz needs to make his argument with much more specific examples (if his object is to show that the regulations are themselves reasons for action). I suggest that if we looked closely, what we would see in any such specific example is that even though there might be ways (many ways) in which the authoritative directive could change the subject’s reasons for acting, we would never be forced to explain this by claiming that the directive was a reason for action in itself. In short, the distinction I used to respond to Raz’s treatment of coordination problems (between the directive’s changing reasons and its being a reason) remains the crucial distinction, on which I think any specific example will founder.

Finally, Raz’s last case is the case of prisoner’s dilemmas. For convenience, let us designate the two choices in the dilemmatic situation as the “selfish” and the “unselfish” choices. (Dilemmas arise out of various situations and various theories about reasons for action, and this pair of labels is not equally apt descriptively in all cases; but the intended reference of the labels is sufficiently clear in all cases, just because every dilemma we are interested in is structurally analogous to the classic prisoner’s dilemma.) Now, on the most common understanding of the classic prisoner’s dilemma, the selfish choice is immoral and the unselfish choice is the only moral choice, quite apart from any intervention by authority. If government steps in to enforce the unselfish choice, it only directs the subject to do what she had overriding pre-existing reason to do anyway. There is no need to appeal to the authoritative directive as a new reason to act to explain how government “solves” this dilemma. The same goes, of course, for all the immense range of analogues of the classic dilemma in which we think the unselfish choice is morally required.

Instead of thinking simply that the unselfish choice is morally required, we might think that the unselfish choice is morally required of each agent on condition that the others are choosing it also. This turns the situation into a sort of coordination problem (to which Raz analogizes prisoner’s dilemmas). We already know that we can account for authority’s role in solving coordination problems without supposing that authoritative utterances are reasons for action in themselves.

Raz also mentions, citing Derek Parfit, that “many moral theories may land their adherents in Prisoner’s Dilemma type situations. The problem does not arise merely through lack of moral fibre.” Id. at 51. Now, Parfit is very clear that dilemmas arise from moral theories only if the theories are agent-relative, that is, if they assign to different agents different goals. D. PARFIT, REASONS AND PERSONS 54-55, 93, 95-98 (1984); see D. REGAN, supra note 42, at 54-65. If, as I have suggested, Raz is at bottom a believer in agent-neutral morality, see supra text accompanying notes 45-46, he need not concern himself with Parfitian dilemmas. That is just as well because I do not think authority can solve a Parfitian dilemma in the relevant sense of “solve.” What an authority with coercive power indubitably can do is to threaten punishment for people who make the selfish choice; this threat, if strong enough, will induce unselfish behavior; and on the behavioral level the dilemma will be solved. But remember, the question we are interested in is whether the authoritative directive constitutes a reason for doing the unselfish act. In one sense, of course, the directive does create a new reason, namely, the punishment that is threatened if the subject does the selfish act. But this plainly does not establish the authoritative utterance as a reason in itself — the reason is, as we
most striking conclusion about political authority—that the scope of government’s authority is likely to vary from citizen to citizen—merely serves to emphasize that political authority is \textit{not} in a class by itself.

Having gone this far, Raz introduces two new, closely related arguments for the bindingness of political authority, the argument from consent and the argument from respect for law. Actually, there is nothing about these arguments, as Raz develops them, that makes them inapplicable in principle to the justification of nonpolitical authority. But he discusses them primarily in connection with political authority; I shall simply ignore the possibility of other applications.

The main idea common to these new arguments is that a citizen may put herself under a more extensive obligation to obey the law of her country than can be accounted for by the normal justification thesis. (Note, the claim is that a citizen may, not that every citizen does.) Consent is a completely voluntary act of submission to authority; an attitude of respect for law is a “semi-voluntary performative submission.”\textsuperscript{80} Neither consent nor respect can create an obligation to obey an essentially unreliable authority; but either consent or respect may expand a citizen’s obligation so that she is bound, at least prima facie, by \textit{all} the commands of a generally reliable authority, including those concerning subject matters where the normal justification thesis might fail to justify

\textsuperscript{80} J. \textsc{Raz}, \textit{supra} note 1, at 98.
for this particular citizen an obligation to obey this particular government. 81

Obviously the expanded obligation Raz attempts to justify by reference to consent or respect for law is much more like the obligation to obey the law that most people believe in than is the limited and patchwork obligation Raz had previously put forward. Also, if the arguments from consent and respect actually work, the language of "obedience" seems much more apposite here than in connection with the earlier, limited obligation. (Indeed, I wonder if Raz's inclination to talk about obedience throughout is not in part a reading back from the specific context to the more general.)

What are the arguments from consent and respect for law, and do they work? First, let us be clear what the arguments are supposed to do. Many philosophers who base political obligation on consent have taken it for granted that consent can create obligation, and have concentrated on trying to find ways around the obvious fact that most people in most societies never give their consent to political authority in any normal sense of "consent." This is what leads to doctrines of tacit consent, and hypothetical consent, and the like. Raz has no truck with this enterprise. Raz sees that many people never consent to the authority of law (and never manifest the attitude of respect for law as he defines it), and he sees also that the important issue in connection with consent is about the consequences of consent, or respect for law, for those who do. What Raz is trying to explain is how consent, or an attitude of respect, can create obligations in the first place.

Let us start with consent. Raz offers both an "instrumental" and a "noninstrumental" argument for the obligation-creating power (or "validity" or "bindingness") of consent. I shall ignore the instrumental argument, because even Raz does not think this argument allows consent to expand the scope of obligation. 82 (It allows consent to create a new motivation for satisfying obligations already explicable by reference to the normal justification thesis, but this is a matter of less interest to us here and now.) What we are interested in, then, is the noninstrumental argument. This is the argument that is supposed to allow consent to expand the scope of obligation. It goes as follows: If one lives in a reasonably just society, then an attitude of identification with that society is intrinsically valuable. Furthermore, the expression of such an attitude of identification is intrinsically valuable. One way of expressing such an

81. Id. at 92, 98.
82. Id. at 90.
attitude, if consent is binding, is to consent to the laws of the society. So, if consent is binding, its bindingness makes available a particular valuable act, namely the expressive act of consent. Therefore, consent is binding.\textsuperscript{83}

In connection with respect for law, there are again two arguments, the instrumental and the noninstrumental, of which only the latter concerns us. First, we should note that what Raz means by respect for law is a belief in one's obligation to obey the law of one's society, which belief is founded on an attitude of identification with one's society. (Remember this is a "semi-voluntary performative submission."\textsuperscript{84}) The argument for the actual bindingness of respect for law seems to be this: If one's society is reasonably just, then an attitude of identification with one's society is intrinsically valuable. The belief in an obligation to obey the law, if it is grounded in this attitude of identification, is an expression of the attitude, and is also intrinsically valuable. Therefore, people who have the belief have the obligation they believe themselves to have.\textsuperscript{85}

So far as I can see, neither of these arguments is valid. Raz does not establish either that consent or that respect for law actually creates an obligation to obey the law. But before I discuss what is wrong with the arguments, let me pause to note what is right in the premises. I fully agree with Raz that an attitude of identification with one's reasonably just society may be intrinsically valuable. I agree further that actions expressing an intrinsically valuable attitude are themselves intrinsically valuable. It follows that it may be valuable to announce one's intention to follow the laws of one's society, even beyond what is required by the normal justification thesis. (This is as close as one can come, in my opinion, to consenting to be bound by those laws.) More often relevant in practice is the fact that it may be valuable to actually follow the laws of one's society, in cases where that is not required by the normal justification thesis, if one follows them as a means of expressing one's identification with the society. All of these claims are important, and Raz is one of the few people who have given them the prominence they deserve.\textsuperscript{86} Nonetheless, these claims do not add up to an argument for an obligation to obey the law, even in those who try to consent, or who have the attitude of respect.

\textsuperscript{83} Id. at 91-93.
\textsuperscript{84} Id. at 98.
\textsuperscript{85} Id. at 98-99.
\textsuperscript{86} Another person who has given prominence to similar claims is P. SOPER, A THEORY OF LAW (1984).
If we focus just on respect, which is more important in this context, what Raz's observations show is that people who have the attitude of respect have a reason for following the law (a reason for action in the strict sense, I might add) which is additional to any reasons suggested by the normal justification thesis. But to have an additional reason is just that, to have an additional reason. It is not to have any special sort of reason which qualifies for the honorific "obligation." In particular, the reason for following the law which flows from an attitude of respect for law is not in any sense an "exclusionary" reason.

Notice that what I have just been denying is the presence of obligation. I suggested earlier that the language of obedience seems somewhat more apt when we are talking about consent or respect for law than when we are justifying authority by the normal justification thesis. I am willing to admit that, as a matter of ordinary language, it may seem proper to say that if I follow a law specifically out of an attitude of respect for the law, then I have "obeyed" the law. But granting this, what follows is that an attitude of respect for law may create (or constitute) a reason for obedience. It does not follow that it creates an obligation to obey. If someone now objects that reasons for obedience only come in the form of obligations to obey—that the notion of obligation is inseparable from the notion of obedience—I can only say that this does not seem so to me. Certainly it does not follow from the particular suggestion about the ordinary language usage of obedience that I made a moment ago. Indeed, if it were true, then I would take it as a conclusive argument against going even as far as I said I was willing to go in conceding that respect for law may convert law-following into obedience. (Perhaps it is also worth restating a different point. Even if law-following based in respect for law does amount to obedience, it does not follow that law-following based on the recognition of law as authoritative under the normal justification thesis is obedience.)

As to why Raz's arguments for the bindingness of consent or respect for law do not work, there is not much to say. The form of the argument about consent is: "If consent were binding, then people would have the

---

87. I confess that in denying there is an "obligation" here, I do not know exactly what I am denying. Many people, including Raz, talk about a special category of oughts known as obligations. I have never felt that I really understand what they are talking about. What I can say — in full confidence that I understand sufficiently well what I am saying — is that when I look where others point and say "obligation," all I see are (moral) reasons which I can make sense of without any such concept. They may be reasons in the strict sense, or they may be "indicative reasons," which may or may not ground indicator-rules. *Cf. supra* notes 40-43 and accompanying text. But they are not, so far as I can understand others' use of the term, "obligations." *Cf. supra* note 43.
opportunity for certain intrinsically valuable acts of consenting. That would be a good thing. Therefore consent is binding.” The form of the argument about respect is: “People who believe, on the basis of an attitude of identification with their society, that they have an obligation to obey the law, have an intrinsically valuable belief. Therefore, they have the obligation they believe in.” But neither of these argument forms is valid. It simply does not follow from “It would be a good thing if X” that “X.” (And therefore, it does not follow from “It would be a good thing if consent were binding” that consent is binding.) Nor does it follow from “People believe X, and it is a good thing that they do so” that “X.” (And, therefore, it does not follow from the fact that some people believe in an obligation to obey the law, and that their having that belief is a good thing, that they have the obligation.)

My perplexity in the face of arguments that seem to me obviously inadequate is only increased by the fact that Raz is not alone in making arguments like this. Many people argue for the bindinguess of promises by saying, in effect, “It would be a good thing if promises were binding. Therefore they are.” Now, if the argument were, that it would be a good thing for us to treat promises as binding, and therefore that we should so treat them, I would have no difficulty understanding that. I could even endorse the argument and the conclusion, subject, of course, to a good deal of clarification about just what is meant by “treating promises as binding.” But what the argument, with the appropriate clarification, would prove, is just that we ought to have an indicator-rule of some sort requiring the keeping of promises. It would not prove that having made a promise is a reason for action in the strict sense. That is what people are normally claiming when they say promises are binding, and it is as an argument to that conclusion that I cannot see any force in claims of the form “It would be a good thing if . . . .” 88

Perhaps the point is this. No one could believe that, in general, we can infer “X” from “It would be a good thing if X.” But some people might believe we can make this inference (or even more plausibly, the inference from “People believe that X, and their having that belief is a good thing” to “X”), when “X” represents a moral proposition. The underlying idea would be that morality is not a set of facts about the universe, but rather a set of ideas and practices we invent. So, if we have

88. To speak of the promise-keeping rule as a mere indicator-rule fails to do justice to the function of promises as devices for coordination. But I am not attempting here a full explication of promising. We have seen in Section I.C., in a different context, that achieving coordination does not depend on authoritative utterances’ being reasons for action in the strict sense. The same conclusion holds in connection with promising.
a moral belief, and if it is a good thing that we have that moral belief, that is all the warrant one could possibly want for saying the moral belief is true. I shall refer to this view about the nature of morality as "conventionalism."

One possible explanation for Raz's arguments about consent and respect for law is that he is a conventionalist about morality. But in fact, I think it is tolerably clear that Raz is not a conventionalist in any ordinary sense. Raz does not make his views on the foundations of morality as clear as one would wish. There is no straightforward discussion of such questions, and at some points Raz seems almost as ambivalent in his relation to conventionalism as in his relation to consequentialism. There are paragraphs where a certain amount of exegesis is required to dispel the impression that there are conventionalist claims and realist claims in neighboring sentences. But at the point where it might seem to the unwary reader that Raz comes closest to endorsing conventionalism definitively, he is completely explicit in the other direction. Having stated the thesis that valuable individual activities must be based on existing social forms, Raz pause to emphasize that this "is not a conventionalist thesis."

If Raz is not a conventionalist about morality in general, there remains a sense in which he accepts a limited and personalized conventionalism. Raz thinks that if I make a commitment to pursue a particular valuable activity (where the value of the activity is not purely a matter of convention), then my commitment, once made, becomes in itself a new reason for pursuing that particular valuable activity instead of others. Because the activity is valuable, I had some reason to pursue it even before I made the commitment. But having made the commitment, I have a new reason, the commitment itself, which is additional to the reason(s) I had before. This idea Raz refers to frequently as the idea that people are "part authors of their own moral world."

It is easy to overlook the appearance of this idea in the course of Raz's arguments about consent and respect for law, but it does appear, and it suggests another ground for Raz's conclusions: Consent and respect for law are binding because they represent commitments to particular valuable activities, and because such commitments in general are

89. See, e.g., J. RAZ, supra note 1, at 3, 327, 344.
90. Id. at 310.
91. See, e.g., id. at 386-89.
92. Id. at 98.
93. Id. at 94, 98.
binding. In fact, this is the best face that can be put on Raz’s arguments. But I think the arguments still fail, because the premise, that commitments are binding, is not true. Here, as elsewhere, Raz has put his finger on an important moral phenomenon—what I shall refer to later as the “stickiness” of commitments—but Raz tries to push it too far. As I shall explain later, there are ways to account for the stickiness of commitments without assuming that commitments are binding; and the only arguments I can find in Raz for the bindingness of commitments turn out to be of the unacceptable form “It would be a good thing if...”

E. AUTHORITY AND THE SPECTER OF MAJORITARIAN PERFECTIONISM

Let me conclude the discussion of authority with a word about the connection between Raz’s chapters on authority and the rest of the book. It is striking that once Raz has finished with the main discussion of authority, he hardly mentions it again. He refers to it very occasionally. But there is really only one place where the discussion of authority seems to do any serious further work. This one place, however, is very important. Raz’s overall political theory is perfectionist. I have previously quoted Raz to the effect that government should promote the moral quality of the life of its citizens. One of the most frequently encountered objections to the pursuit of perfectionism by a democratic government embodying the principle of majority rule (even if subject to constitutional limitations) is that under such a scheme the majority gets

94. See infra text accompanying notes 165-70.
95. Let me add two final points. First someone might suggest that Raz is not a conventionalist, but he is a rule-consequentialist, and that justifies his arguments. I cannot pause to discuss this possibility at length, but I remind the reader that rule-consequentialism has its own ambiguous relationship to conventionalism. If rule-consequentialism bases itself on actually existing social rules, provided they meet some threshold of adequacy, then it takes large steps towards conventionalism. On the other hand, it is not clear that an “ideal” rule-consequentialism (not based on conventional rules) has ever been coherently stated. (For argument relevant to this point, though not a general proof, see D. REGAN, supra note 42, at 83-104.) Second, at one point Raz defends himself against the claim that his notions of consent and respect for law involve a vicious self-reference. J. RAZ, supra note 1, at 99. Raz partly defines consent, which creates obligation, by the fact that it is intended to. Although some may see such a definition as objectionable, I agree with Raz that it is not. I want to emphasize that my problem with Raz’s arguments is not the self-referential quality of consent or respect for law. Even granting that the self-reference is not a problem in itself, I just do not see any good argument that consent or respect for law are binding. The self-reference, even if it is non-vicious, cannot supply the argument. We cannot explain why consent is binding just by pointing out that it is intended to be.
96. Id. at 262, 427.
97. Id. at 415.
to impose its view of the good life (within limits, perhaps) on the minority. The majority view is imposed just because it is the majority view, and that is inappropriate.

As Raz points out, this objection is confused, in a way that a proper conception of authority can help to dispel.98 The majority, if they are behaving as they should, do not impose their view just on the ground that it is their view. They impose it on the ground that, so far as they can see, it is the right view, and is therefore morally binding on the minority by its own force. (This follows from the dependence thesis about authority.) More importantly still, the whole structure which puts such authority in a majoritarian government is justified only if most people will in general behave more morally by following the commands of the government than by trying to make all decisions on their own. (This is the normal justification thesis.)

One might argue against Raz that a majority will be wrong on questions about what makes a morally good life for individuals more often than not. Or, one might argue that it is never useful on balance to base any governmental decision on views about the good life, because of the inevitable overgeneralization. Better to leave all such decisions to individuals. Raz's answer would be that these are partly empirical claims about how majorities will do at making certain decisions (including decisions about which decisions to make at a governmental level). Majoritarian governments simply do not have authority over questions about the good life if these claims are true. But Raz thinks the claims are overstated, at least for some societies. And so do I. In any event, once we understand how authority is limited by the dependence thesis (insofar as it is conscientiously exercised), and by the normal justification thesis, we see that there is no reason of principle why government should not concern itself with encouraging the good life. And there is no reason of principle why the views of majorities on such questions, even though they may sometimes be mistaken, should not be recognized as authoritative.

II. VALUE

As I originally conceived this half of the essay, there was to be a substantial discussion of incommensurability flanked by brief comments on well-being, social forms, and autonomy. The flanking sections grew beyond my anticipation—I hope not beyond the point of diminishing

98. Id. at 157-62.
interest. But the section on incommensurability remains the most important so far as the differences between Raz and myself are concerned.

A. WELL-BEING

Raz treats it as axiomatic that well-being is central. He writes, early in the portion of the book that sets forth his own substantive views: "Morality is thought to be concerned with the advancement of the well-being of individuals." The explicit claim here is only that morality "is thought" to be concerned with individual well-being, but it seems clear that Raz himself is among those who think this.

Raz has many interesting things to say about the nature of well-being, which I shall summarize crudely as follows: An individual's well-being is determined by the degree to which she succeeds in achieving her genuinely valuable comprehensive goals over a lifetime.

One of Raz's main objects in discussing well-being is to distinguish it from self-interest. To summarize crudely once again, self-interest differs from well-being in that self-interest focuses more on the satisfaction of biological needs and on contentment. Thus, the person who willingly shortens her own life by working to ease the sufferings of people afflicted with some highly contagious disease acts against her self-interest even as she promotes her well-being. Conversely, the person who gets great pleasure out of success in a project which she mistakenly believes is genuinely valuable acts in accord with her self-interest (to which her mistake about the value of the project does not matter) but does not increase her well-being (to which the mistake does matter). Raz eventually concludes, "[People's] well-being, and not their self-interest, matters most both morally and to people themselves."

So far I agree completely. If we must ground morality on the promotion of either people's well-being or their self-interest, well-being wins hands down. But I think Raz does not go far enough. In the end, well-being is not what matters any more than self-interest is. What does matter ultimately is just those valuable projects that, in Raz's scheme, contribute to well-being. And they matter, not because they contribute to anyone's well-being, but because they are valuable in themselves.

99. Id. at 267.
100. See id. at 288-99. For further comments on the meaning and importance of "success," see infra text accompanying notes 104-05, 144-47, 152-53, 187-89.
101. See J. RAZ, supra note 1, at 317.
In this essay I cannot begin to give a full argument against the claim that well-being matters. I would not even pause to controvert an assumption so in tune with conventional thinking, were it not that there is a striking ad hominem argument against Raz’s taking well-being as fundamental. Let me emphasize, the point here is not that Raz makes worse arguments than most people. Rather, he makes better arguments than most people, which bring him uncommonly near the truth; but he fails to take the very last step.

Raz thinks we should not try to exclude ideals from politics. In the chapter entitled “The Exclusion of Ideals,” Raz argues that if we admit the existence of valid moral ideals, we cannot defend a politics of mere desire-satisfaction. The argument, which I fully endorse, goes as follows: People regard themselves as having reasons for their desires. (Perhaps we should say “for most of their desires, including the most important ones.”) Specifically, they regard themselves as desiring the things they desire because those things are actually valuable, independently of their being desired. Furthermore, people’s desires are conditional on the correctness of their judgments of value. I cannot, at this moment, recognize any of the value judgments I make at this moment as being mistaken, but I can recognize with no difficulty that any of them might be mistaken and that some of them probably are. Other things being equal, I would prefer that those desires of mine which are based on mistaken value judgments not be satisfied. In sum, people do not want their desires to be satisfied just as their desires. But if people do not desire the satisfaction of their desires as such, there is no reason for government to aim at the satisfaction of desires as such. In short, politics should not aim at mere desire-satisfaction.

I agree with this argument right down the line. All I would add is that it also undermines politics based on maximizing well-being. Individual agents do not aim at their own well-being. They have projects, and success in their projects contributes to their well-being if the projects are valuable. But still, they do not aim at their well-being. Raz understands this (and that is why I can offer this argument ad hologem). “[I]t is neither tautological nor true that we pursue our well-being.” And, “it is false that we pursue our goals because their pursuit serves our well-being.”

It remains only to observe: If the reason it is inappropriate for politics to focus on desire-satisfaction is that the individual owners of the desires are not interested in the satisfaction of their desires as such, then it seems

102. Id. at 316.
103. Id. at 317.
equally inappropriate for politics to focus on well-being, if that is not what the individuals whose well-being is in issue are aiming at.

There is, of course, an important difference between the case of desire-satisfaction and the case of well-being. Desires may be based on mistaken value judgments. If we promote desire-satisfaction without looking behind the desires, we may put government resources behind desires whose satisfaction is not genuinely valuable at all (and which the possessors of the desires therefore do not really want to be satisfied). In contrast, well-being is *defined* by Raz in terms of success at valuable goals, so there is no danger that in promoting well-being we will spend resources on goals of no genuine value, or goals which the possessors of the goals would abandon if they were disabused of some mistake.

This difference between promoting desire-satisfaction and promoting well-being is real enough. But to my mind it does not undermine the basic point: Something has gone wrong when government aims at promoting, on behalf of individuals, something other than what the individuals conceive of themselves as aiming at. In short, something has gone wrong if government aims at promoting well-being.

The main point I wanted to make in this section is now established. But someone might still ask: Granting that something has gone wrong if government aims at promoting well-being, has anything gone wrong that matters? If the nature of well-being is such that promoting well-being necessarily involves promoting goals of genuine value, does it make any difference in practice whether agents or governments aim at promoting well-being or at promoting particular valuable goals?

The most important part of the answer to this question is that, whatever may be the case with individual agents, it very definitely does matter whether government aims at promoting well-being or at promoting particular valuable goals. In order to emphasize that that is the most important part of the answer, I am tempted just to concede for purposes of argument that it makes no difference whether each individual agent aims at his own well-being or at his particular goals, and to go straight to the question about government. Tempted, but not determined. Even the case of the individual agent is more interesting than it might appear, and I shall pause over it briefly.

I am prepared to concede this: If we first specify the agent's goals fully and carefully, then it makes no difference to his behavior whether he aims at his goals or at his well-being. Nonetheless, it may make a
difference to what we as observers think of as the agent's goals (and conceivably even to what the agent himself thinks of as his goals) whether one is inclined to attend mainly to the goals or mainly to the well-being they contribute to.

Let me give an example of a case where I think Raz is led to misconstrue an agent's goals by excessive concern with the agent's well-being. In one and the same paragraph, Raz first says: "People pursue goals in order to succeed in them. Since they care about their success, failure reflects adversely on their well-being, on the success of their life." Then a few sentences further on, we find: "This does not mean that the success of all his projects and goals makes a difference to the agent's well-being. What counts is how successful he was in pursuing them, how he managed in his pursuit of them." What explains this seeming volte-face on the question whether success matters to well-being? The clue, I think, is in the continuation after the second quoted sentence. Raz goes on:

A person who actively campaigned to prevent the use of DDT may see its use actually increase in his lifetime, without this casting a shadow on his well-being. What matters from that point of view is the way he pursued his objectives and the contribution he made, relative to what could be expected, to his cause. It seems to me that, in order to make it more likely that the agent can be counted a success and that we can regard his work against DDT as contributing positively to his well-being, Raz has implicitly reconstrued the agent's goal. He has taken the goal of eliminating DDT and transformed it into the goal of struggling meritoriously for the elimination of DDT.

It is easy enough to sympathize with Raz's motives. For many purposes, what we care about is not people's success, but their efforts. If we are judging the agent's moral character, we will not think it matters whether he was successful, but only how he pursued the campaign. Furthermore, we think it a good thing if people can take satisfaction in good work done in support of a losing cause. We may therefore hope that people whose goal is to eliminate DDT will at least have as a subsidiary goal the goal of doing well in the campaign to eliminate DDT. That way, even if they fail in their primary goal, they need not fail utterly.

Still, it would be very odd to have as one's primary goal "doing well in the campaign to eliminate DDT." Caring about doing well in the campaign ought to flow from caring in the first instance about the goal the campaign is directed at. It seems to me Raz had it right the first

104. Id. at 298.
105. Id.
time. People care about their goals; and failure to achieve their goals is a cost to their well-being. Whatever the virtue and the subsidiary benefits of good work in a losing cause, we should not be so concerned to make virtue correspond with success and well-being that we reconstrue agents' goals in order to maximize the correspondence.

Now back to the more important aspect of our question. Even granting that it makes no difference whether the individual agent aims at his (precisely formulated) goals or at his well-being, it makes a great difference whether government aims at promoting individuals' goals or at promoting well-being. (More accurately, there is a great difference of principle here, which could well make a difference in practice.) It makes a difference because one of the questions government faces is how it should allocate resources among the projects of different individuals.

Raz plainly wants a substantial element of egalitarianism in the allocation of resources, even if he does not believe in a significant right to equality as such. The egalitarian element is brought into Raz's theory, which focuses on promoting well-being, by the fact that well-being is, for Raz, a diminishing and satiable value.106 (I do not know that Raz ever says this explicitly. He does say generally that "The ideals at the foundation of morality and politics are all diminishing and satiable principles." And he argues specifically that happiness, which seems essentially the same as well-being, is diminishing and satiable.107) Because well-being is diminishing and satiable, there is a diminishing (and eventually zero) marginal return from social resources spent on promoting any individual's well-being. If society's goal is to promote well-being, there is a resulting tendency to egalitarianism.

We do not get the same tendency if we focus on promoting individuals' goals. Suppose we begin by supporting some goal of Jones's. We may suppose that the value represented by that goal is diminishing and satiable. Even so, once we have fully supported that goal, or supported it as far as seems reasonable, there is no reason why we should not choose as the next goal to support another goal of Jones's (instead of some goal of Smith's). Once we have supported Jones's second goal to satiation or seriously diminishing returns, we can go on to Jones's third goal, still ignoring Smith, if all we are concerned about is individuals' goals considered just as goals. At some point we will exhaust Jones's goals, or (probably sooner) Jones. But it should be clear that the egalitarian tendency is

---

106. For the meanings of "diminishing" and "satiable" see supra text accompanying note 24. 107. J. Raz, supra note 1, at 241.
likely to come into play later if we focus on individuals' goals than if we focus on well-being.

Of course I do not mean to deny that society might have the basic object of promoting individuals' goals and might constrain pursuit of that object by an independent principle that we ought to promote different individuals' goals equally. That is plainly a possibility. My point is simply that there is not the same egalitarian tendency built into the idea of promoting goals that there is built into the idea of promoting well-being (as Raz interprets it). The basic reason for the difference is that a goal is a goal, in whatever life it occurs; and we can promote goals without paying attention to how they are distributed over individual lives. (Of course every goal must be some individual's goal, but even so, we can promote goals without regarding it as significant whose they are.) The notion of well-being, in contrast, focuses our attention precisely on the boundaries between lives. How much the fulfillment of a goal contributes to the well-being of the individual whose goal it is depends (because well-being is diminishing and satiable) on how other goals are faring which occur within the boundaries of the same individual's life.

The topic of the difference between pursuing or promoting goals and pursuing or promoting well-being is too deep for a complete discussion here, though I think everything I have said is correct as far as it goes and also that nothing I have said is misleading. Someone might want to argue that even promoting goals leads us to attend with particularity to whose goals they are, on the ground that when Jones does physics (for example), his goal is not just "that physics should be done," but rather "that I, Jones, should do physics." So, it is not just that every goal is someone's goal; rather, every goal contains its owner's identity as part of its essential description.

For myself, I do not think this is true. I think Jones's goal really is in its essence "that physics should be done" rather than "that I, Jones, should do physics." (It is an accidental feature of Jones's goal that, seeing no better way to advance physics or to spend himself, he has decided to do physics himself.) If Jones chooses his goals on the basis of their value, how could his goal be essentially that he do physics? How could he believe that his doing physics is specially valuable because it is he who does it?

Even if I am wrong about this, I doubt that it matters to the question of allocation of resources. Even if each of Jones's goals comes stamped with the name "Jones" in its essence (and similarly for Smith's goals), there is still nothing in that which tends to force government,
after it has supported one goal of Jones's, to turn its attention to a goal of Smith's. It remains true that a goal is a goal is a goal, whatever life it occurs in and whatever name comes stamped on it. Only when government turns its attention to well-being does the connection between particular goals and particular individuals matter to the selection of goals to be supported.

In sum, it does make a difference whether government focuses on promoting goals or on promoting well-being. The latter, but not the former, has a powerful built-in tendency to equality in resource allocation. Unfortunately, as I have pointed out, the fact that individuals aim at their goals, and not at their well-being, seems to be an argument for government’s promoting goals instead of well-being. What, then, of equality?

Someone might argue (Raz might even argue) that we want the built-in egalitarian tendency of the focus on well-being, and that that is reason enough to say government should focus on well-being even though well-being is not precisely what the individual beneficiaries of government's action have as their aim. Alternatively, we might say government should focus on promoting individuals' goals, but then reintroduce a separate equality constraint in some form. (This, I take it, Raz would reject.)

For myself, I would reject both a foundational role for well-being and an independent equality constraint, for which I simply cannot see any adequate grounding. (Remember, all I am discussing here is equality in resource allocation.) I can do no more than hint at what I think is the right place to look for such egalitarian tendencies as may exist in a correct political theory: In my view, community is an objective, intrinsic good, something valuable in itself (and not merely valuable for the members of the community), which contributes to the value of the universe in which it occurs. By "community" I mean something like friendship writ large, insofar as that is possible over a polity. I suspect that the degree of equality in the distribution of wealth and income is positively correlated with the achievement of the good of community (at least as an empirical matter and possibly as a conceptual matter also). To the extent that equality matters, that, I think is how.

Let me tack on one last observation, about a paradox Raz raises and, to my mind, inadequately resolves. In the passage which I have already adverted to where he discusses the satiability of happiness, Raz suggests that many people have failed to accept that happiness is satiable because they worry that "there would then be no reason for giving the
perfectly happy person further pleasure or rewarding experiences."\(^{108}\) Actually, this conclusion is correct on Raz's own premises. If the point of morality is to promote well-being; if well-being is essentially equivalent to happiness (as I think we may assume); and if the happy person's incremental success with his projects does not make him happier (as Raz asserts); then we have in fact no reason to help him towards this incremental success. Raz tries to avoid the conclusion. "A perfectly happy person has ambitions, pursuits, hobbies, etc. Like everyone else, he aspires to success in what he does."\(^{109}\) This is true enough. But it is irrelevant if further success does not increase happiness and if happiness is what matters. Raz must be assuming implicitly that happiness is not all that matters after all. Success in valuable projects matters independently of its contribution to happiness. In sum, we cannot explain why we have reason to promote the incremental projects of the perfectly happy person unless we recognize that what we ought to promote in others' lives is the achievement of their (valuable) projects, not their happiness or well-being.

**B. SOCIAL FORMS**

Raz makes much of what he calls "social forms" (what others might call "social practices" or "conventions"). Social forms are important in large measure because they underwrite the possibility of "comprehensive goals." A goal of mine is a comprehensive goal if it has "ramifications which pervade important dimensions of my life."\(^{110}\) Among other claims, Raz attempts to defend the thesis that "a person can have a comprehensive goal only if it is based on existing social forms, i.e., on forms of behaviour which are in fact widely practised in his society."\(^{111}\) Although I agree with much of what Raz says about social forms, I have doubts about this strong thesis, as I shall explain in due course.

It is certainly true that many of the comprehensive goals one can have in our society, one could not have outside society or in a society different in relevant respects. "One cannot pursue a legal career except in a society governed by law, one cannot practise medicine except in a society in which such a practise is recognized."\(^{112}\) As Raz goes on to explain, one might be born into a society which hitherto knew nothing of medicine with an innate knowledge of how to cure disease; and one

---

108. *Id.* at 243.
109. *Id.*
110. *Id.* at 308.
111. *Id.*
112. *Id.* at 310.
might then cure many sick people; but one would still not be a doctor with all that means in our society. Much the same can be said about marriage, to take an example from the sphere of personal relations. There could be long-lasting cohabitation of individual men with individual women in a society which had no institution of marriage. But there could not be the rich relationship of marriage as we know it, not even if we restrict our view to the lives of the two parties.

Raz summarizes the matter thus: "[E]ngaging in the same activities will play a different role, have a different significance in the life of the individual depending on social practices and attitudes to such activities." I would put the same basic point somewhat differently. I would say that engaging in the same behavior will have a different significance depending on the social context, and the reason is that the same behavior will constitute different activities. Curing disease in a society without medicine and being a doctor in a society with medicine may often involve the same behavior. But even when they do, they are different activities. Their nature and identity as activities depends on the full mental state of the person engaged in them, which will be very different in the differing contexts. The full mental state of a doctor cannot exist in a society where doctors are unknown.

Now, the main thing I want to emphasize, in my brief discussion of social forms, is a point that Raz himself states unequivocally (but that he also seems occasionally to forget): Nothing said so far about social forms entails a conventionalist theory of value. The thesis Raz is defending "is not a conventionalist thesis." "It does not claim that whatever is practiced with social approval is for that reason valuable." Nor, I might add, does it claim that whatever is practiced without social approval is for that reason not valuable.

Raz's thesis, in other words, is about the availability of certain activities to individuals. It is not a thesis about the source of the value (or disvalue) of such activities as may happen to be available.

Notice this: Even though value is not conventional, the social facts which determine the availability of activities are very important in determining the value achieved by various behaviors. The value of an activity depends on what activity it is. What activity some behavior constitutes typically depends on the social context in which it occurs. Therefore, the

113. Id. at 311.
114. Id. at 310.
115. Cf. id. at 309 (Comprehensive goals may be "based on a shared perception of a social form while deviating from it in some respects.").
value of behavior typically depends on the social context in which it occurs. But that still does not mean that the value of various behaviors or activities is determined, except incidentally, by social valuations of them. There can be completely objective principles about what activities have what values, despite the fact that social context plays a major role in making the behaviors and the activities what they are.

Raz never even hints at anything inconsistent with this in his principal discussion of social forms. It is when he talks about social forms in other parts of the book that he tends to slide towards conventionalism. For example, in the chapter on the incommensurability of values (which I discuss more fully in the next section) Raz says, "The dependence of value on social forms makes the brute existence of incomparabilities no more than one would expect." But if value depends on social forms in the way we have so far described, it is not at all obvious why this dependence should lead us to expect incomparability. Most particularly, it would not lead us to expect incomparability on the straightforward conventionalist ground (which Raz seems to appeal to later in the same paragraph) that social conventions about the value of various behaviors and activities are finite and may just run out. As we have seen, social conventions about value are not the standards by which the value of various activities are determined. Social valuations partly constitute the social conventions, and they are therefore part of the context which determines the true value of various activities. But the role they play is such that there is no reason to think objective valuations must run out, or even tend to run out, when social valuations do.

(A somewhat more plausible argument from the importance of social forms to incommensurability might be the following: Social forms are unlimited in their possible variety, and it seems hard to believe that the universe of objective value, if it exists, includes standards appropriate for evaluating all the social practices the corporate imagination of humankind can devise. To my mind, this is still not very persuasive. Even wildly varying social practices may have such value as they have because of their content as considered under a small number of headings, such as the amount of intrinsically valuable knowledge called upon by participation in the practice, the nature and degree of identification with other people manifested in participation in the practice, and so on. If this is the case, a virtually unlimited range of social practices could be evaluated by a relatively compact set of objective standards.)

116. Id. at 344.
I have now dealt with the aspects of social forms that most concern me—both the aspect of Raz's treatment that I most want to agree with (about how social forms condition the significance, and thus the value, of behavior) and the aspect I most want to disagree with (linking up social forms and incommensurability).

At the risk of dealing with too many topics too superficially, let me comment on another use to which Raz puts his arguments about social forms, which is very important to Raz. I summarized much earlier Raz's argument that political theory cannot be founded on individual rights to autonomy. The argument goes as follows: (1) Autonomy requires having a range of valuable comprehensive goals available to choose from. (2) Comprehensive goals must be based on existing social forms. Therefore, (3) autonomy requires the existence of social forms. But, (4) the nature of rights is such that no individual can have a right to the existence of a social form. Therefore, (5) no individual can have a right to autonomy.117

I have doubts about Raz's notion of autonomy, which I shall discuss in the last section of this essay. But for the moment let us grant proposition (1). Proposition (2) is the thesis I quoted at the beginning of this section. Although I fully agree with Raz that the most valuable comprehensive goals available to people in a society like ours are based on social forms, I am not sure Raz ever proves that all comprehensive goals must be based on social forms. Raz points out that one could not be a bird watcher (with all that means to us), nor could one be married, in a society that recognized neither bird-watching nor marriage. But even in such a society, one could devote much of oneself to learning what one could about birds, and one could have a continuing and moderately complex monogamous heterosexual relationship. Very likely one's watching of birds would not be as valuable as bird-watching, and one's monogamous relationship might not be as valuable as marriage. But it requires more argument than Raz gives to show they could not be sufficiently engaging and valuable to be comprehensive goals.

Actually, Raz's main argument for the necessity of social forms is an argument I have not yet mentioned. Raz says that an individual cannot acquire a comprehensive goal by explicit deliberation. Rather, such a

---

117. For a comment on the issue of whether Raz really makes this argument, seeinfra text accompanying notes 127-31.
goal can be acquired only by habituation.\textsuperscript{118} I think there is a large dollop of truth in this claim.\textsuperscript{119} But why could not the non-socially-supported watcher-of-birds acquire his comprehensive goal of knowing as much as he can about birds by habituation which is entirely of his own doing? It takes time, but I see no reason why it necessarily requires a social form. Similarly, the participants in the continuing monogamous relationship which is not marriage must give it some time before it is a relationship at all. But why should not they succeed in habituating themselves and creating a relationship (along with comprehensive goals for each of them) that did not exist before and that society does not contribute to? Perhaps no really complex comprehensive goal could be invented by one or two individuals. We may all need the help of society to initiate us into any pursuits as valuable as the most valuable pursuits that are actually available to us in the world as it is. But that is not the claim Raz started out with, which was that all comprehensive goals require existing social forms.

If that strong claim must be abandoned, Raz can argue that social forms are required by the principle of autonomy only by claiming either (a) that the minimum range of choice of comprehensive goals that counts as autonomy is broader than the range of goals individuals could construct for themselves, or (b) that the principle of autonomy requires government not only to guarantee people autonomy but to give them as many choices as possible. Either or both of these claims may be defensible. But reliance on either would complicate Raz’s argument against the rights-theorist.

There is a final problem with Raz’s argument worth mentioning. Even if we grant that autonomy requires the existence of social forms, is it true that no individual can have a right to the existence of a social form? According to Raz, one has a right only if an aspect of one’s well-being is, other things being equal, “a sufficient reason for holding some other person(s) to be under a duty.”\textsuperscript{120} The rights-theorist might object that this is a rather tendentious definition, especially when he sees what Raz does with it, but let that pass for the moment. Raz denies that he (or anyone) has a right to the existence of social forms because “[my interest] is not enough by itself to justify holding others to be duty-bound to make

\begin{itemize}
\item \textsuperscript{118} J. Raz, \textit{supra} note 1, at 311.
\item \textsuperscript{119} Cf. infra last paragraph of section II. C. (discussing what counts as making a commitment to playing the violin).
\item \textsuperscript{120} J. Raz, \textit{supra} note 1, at 166.
\end{itemize}
sure that my society shall offer [the opportunities the social forms repre-
sent].”

"The maintenance of a collective good [such as a social form] affects the life and imposes constraints on the activities of the bulk of the population, in matters which deeply affect them. It is difficult to imagine a successful argument imposing a duty to provide a collective good on the ground that it will serve the interests of one individual."

The argument here appears to be that no individual's interest is weighty enough to justify imposing on his fellow citizens the burden of maintaining a social form. Unfortunately, it is not clear that Raz is entitled to raise the issue of weight. Raz recognizes that I have some reason, other things being equal, to do whatever will promote Jones's autonomy. And he tells us that "Duties are not reasons for action of a great weight. They are a special kind of requirement for action . . . ." They are special, if I understand Raz aright, in having peremptory force. Now, the rights-theorist might well insist that the reason I have for promoting Jones's autonomy, however great or small its weight, is a peremptory reason in Raz's sense. It excludes me from asking, for example, whether Jones will make good use of the opportunities I provide for him (just as, if I owe Jones ten dollars, I have a reason to give him ten dollars without considering, at least over a wide range of possibilities, what he will do with it). There are other ways also in which my reason for promoting Jones's autonomy might be thought to be a peremptory reason.

In sum, the rights-theorist might insist that I do have a peremptory reason, and thus a duty (though perhaps a very weak one), to do whatever is necessary to promote Jones's autonomy, including contributing to the existence of social forms. And because this particular duty to promote Jones's autonomy (it may not be the only duty I have to promote Jones's autonomy) depends solely on Jones's interest, it justifies us in attributing a right to Jones. (Raz might respond that any duty I have based solely on Jones's interest will be so weak, or so regularly overridden, that it would merely clutter our moral discourse to insist on its existence. But that response seems to bring in considerations of a new and indefinite kind.)

It may seem that I am logic-chopping, holding Raz to his definitions of "rights" and "duties" with a pedantic strictness. But there is more than a logical problem here. Consider what the rights-theorist might say

121. Id. at 206-07.
122. Id. at 203.
123. Id. at 407.
124. Id. at 195.
in support of his earlier-mentioned claim that Raz's definition of rights is tendentious. I imagine the rights-theorist saying:

Raz himself believes that he has a duty to promote others' autonomy. It is because he has this duty that government may tax him to secure funds for its efforts to promote autonomy.125 This duty is presumably based on the aggregate of other persons' interests in their autonomy. The only reason Raz denies that each other individual has a right to Raz's contribution to that individual's autonomy is that each individual is affected so slightly by those very acts of Raz's that, because of their aggregate effects on all individuals, Raz is duty-bound to do. This seems artificial. The issue of whether there are rights to autonomy should not turn on the accident that any particular act by Raz (even an act he is required to do on behalf of others' autonomy) has diffused effects. When I claim there are rights to autonomy, I do not mean to make a claim about the concentration or diffusion of the effects of Raz's acts, but rather a claim about the nature of the interest(s) that, singly or in the aggregate, require the acts. I mean to say that these interests are of a special character. They are special, for example, in that they require Raz to contribute to the promotion of others' autonomy without stopping to inquire just what use they will make of it. It is because the interests are special in this way that I was able to use Raz's definition of duties as peremptory reasons against him. But the real point is just that these are interests of a special kind. That is what I mean when I call these interests "rights," and it is a claim Raz sidesteps rather than confronts.126

My point in all of this is not that the rights-theorist is right and Raz is wrong. Anyone who has read this far knows my sympathies are much more with Raz. But I do think this argument against the rights-theorist is shaky. (From my perspective, there is an obvious way to answer the rights-theorist, a way which would unfortunately be unacceptable to Raz. If we deny that well-being is the foundation of morality, then we can deny flat out that Raz's duty to promote others' autonomy is grounded on those others' interests as interests. The question of the "special nature" of the interests never arises.)

Raz adopts a somewhat different way of dealing with the rights-theorist in a later chapter. After discussing an objection to his definition of rights which overlaps with the objection I have put in the mouth of the rights-theorist, Raz says,

125. Id. at 415-18.
126. It should be noted that Raz devotes some pages to the claim that rights are based on interests of a special nature. Id. at 187-91. But it does not seem to me he really disposes of the claim. Of course, the claim itself is far from perspicuous.
The argument of Chapter Eight [which is what we have been discussing] did not rest on the definition of rights [to which we have been referring]. It rests on the belief, to be defended in the last part of the book, that personal autonomy depends on the persistence of collective goods, and therefore that the notion of an inherent general conflict between individual freedom and the needs of others is illusory. . . . This fact, rather than any definition, undermines the individualist emphasis on the importance of rights.127

Despite Raz’s disclaimer, it seems to me that the argument of Chapter Eight does depend on the definition of rights. (For example, “It is difficult to imagine a successful argument imposing a duty to provide a collective good on the ground that it will serve the interests of one individual.”128 And, “My interest in being autonomous . . . is not enough by itself to justify holding others to be duty-bound to make sure that my society shall offer [a wide range of social forms].”129) What Raz later implicitly recognizes is that he does not need to show there are no rights to autonomy. All he needs to show is that the significance of autonomy outstrips anything we can capture in talk about rights. Raz makes a strong argument for this claim when he argues that autonomy is valuable as an achievement, not as a capacity.130 This argument is alluded to in Chapter Eight, but Raz specifically says the argument of Chapter Eight does not depend on it.131

It may be just as well that Raz does not need to show there are no rights to autonomy. The truth is he comes perilously close to committing himself to the existence of such rights when he tries to explain in Chapter Fifteen why government may promote autonomy by redistributive taxation.132 Raz argues that failure to promote autonomy would constitute harm, on the ground that one harms someone when one makes him worse off than he is “entitled” to be133 or when one fails in one’s duty “to” him to improve his situation.134 “Entitlements” are very like rights, and duties “to” (as opposed to duties regarding) are often thought to have rights as their correlates.

127. Id. at 250.
128. Id. at 203.
129. Id. at 206-07.
130. See infra text accompanying notes 177-80.
131. J. Raz, supra note 1, at 205.
132. Id. at 415-18.
133. Id. at 414.
134. Id. at 416.
C. INCOMMENSURABILITY

Possibly the most significant difference between Raz and myself is on the question whether there are incommensurable values. Raz thinks there are. Indeed, Raz thinks most of the valuable activities around which one might build a life represent incommensurable values. In contrast, I am inclined to think that all values are commensurable. There is no question of proof here. Raz points out the logical possibility of incommensurability; he points to some features of our experience that he thinks indicate values are incommensurable; and he responds, explicitly or implicitly, to some possible arguments against his position. I, conceding that incommensurable values are logically possible, shall offer similar observations on the other side.135

First, what does it mean to say that values are incommensurable? Two valuable things are incommensurable in respect of their value if (a) neither is more valuable than the other, and (b) they are not of equal value. If neither is more valuable than the other, and if their values are also not equal, then clearly their values are incomparable. They do not appear on any common scale of ordering.136

It may seem to the reader at first that if neither of two values is greater than the other, then the two values must be equal. If this were true, then incommensurability as defined would be impossible. But it is not true, at least not as a matter of the pure logic of comparisons. To

135. No discussion of incommensurability should fail to mention the splendid work of clarification and underbrush-clearing in Griffin, Are There Incommensurable Values?, 7 PHIL. & PUB. AFF. 39-59 (1977). But Raz's arguments for incommensurability are not the arguments Griffin demolishes. And I am unwilling to rely on Griffin's argument in favor of commensurability, which depends too much on the premise that value is a matter of desire-satisfaction.

136. It will occur to the astute reader that I am submerging a potentially important distinction between values being incommensurable and value orderings being incomplete. I do so for a number of reasons. First, Raz talks only about incommensurability, not incompleteness (even though the sort of example he gives to illustrate the possibility of incommensurability is the same sort of example that Parfit gives to illustrate the possibility of incompleteness). D. PARFIT, supra note 79, at 431. Furthermore, I think incommensurability is the more significant possibility in the context of Raz's discussion. If what underlies our “value orderings” is preferences, then it is the most natural thing in the world to suppose that the orderings may just be incomplete. But if we are operating in the world of objective values, as Raz is, then pure incompleteness seems less intuitively plausible (to me, and perhaps to Raz). Moreover, an obvious way exists in which complete incommensurable objective orderings can give rise to an incomplete synoptic orderings. We can compare states of affairs by reference to all the various incommensurable orderings, and then say that one state is superior to another on the synoptic ordering if and only if it is at least as good on every individual ordering and better on at least one. This is a Pareto inspired mode of combination, which obviously produces an ordering, but an incomplete one. So, we can generate incompleteness in a single ordering out of incommensurability among many complete ones. Hereafter, I shall assume that incommensurability is the real issue.
show that incommensurability is possible, Raz invites us to consider an example, of a type that has now become standard. Imagine that I am considering whether to spend the afternoon walking in the park or reading a book at home. By stipulating properly the conditions of the walk (what the weather is like, how crowded the park will be, and so on) and the conditions of the afternoon reading (what book I shall read, whether I shall have a glass of Scotch), we can come up with alternatives such that it seems to me neither is more valuable than the other. (For the present, I am describing the example as the proponent of incommensurability would describe it. I do not mean to concede the absolute accuracy of every element of the description.) Say that I see nothing to choose between a particular walk in the park and reading with a glass of Scotch. So far so good. Now, in order to show that this nonsuperiority of either alternative does not entail that the two are equal, imagine a small improvement in one. Say we substitute for the glass of Scotch a glass of port, which I would slightly prefer. Now, if the original alternatives were actually equal in value, then, given that I prefer the reading-with-port to the reading-with-Scotch, I ought, by transitivity, to prefer the reading-with-port to the walk in the park. But we can perfectly well imagine, without making me seem irrational, that this is not the case. Even though I unambiguously prefer reading-with-port to reading-with-Scotch, I might well find that the choice between reading-with-port and walking in the park poses the same difficulty as the original choice between reading-with-Scotch and walking in the park. If that is so, then the difficulty of choosing between the original alternatives could not have resulted from the fact that they were equal in value. It must have resulted from the fact that they were simply not comparable at all.

It is unclear, both as Raz presents this example and in my recapitulation of it, whether the example is meant to involve mere subjective preferences, or judgments of objective (that is, ultimately non-preference-based) value. If the example is about subjective preferences, it obviously does not prove that non-preference-based (objective) values may be incommensurable; and it is objective values that both Raz and I are ultimately concerned with. Even if we take the judgments in the example to be judgments of objective value, the example is weaker evidence for the existence of incommensurability than it might seem to be at first. But, not wishing to trivialize the subject by concentrating on a trivial example, I shall postpone discussion of this further point. Raz offers the

---

137. Cf. supra text accompanying notes 113-16.
example to clarify the concept, and to illustrate the logical possibility, of incommensurability. For these limited purposes, the example succeeds.

The importance of incommensurability, for Raz, lies not in examples about how to spend an afternoon. As I have mentioned, Raz thinks most of life's major activities are incommensurable one with another. A life as a lawyer is incommensurable with a life as a clarinetist, and with a life as a teacher.138 A career as a graphic designer is incommensurable with a career as a livestock farmer or a gliding instructor.139 And so on.

So, why does Raz think such choices involve incommensurable values? One of his arguments is that people frequently say such values as these are incommensurable, or refuse to make such comparisons.140 I do not find this argument convincing. There are a great many views that people might be expressing when they talk or behave this way, other than the philosophically sophisticated view about incommensurability Raz attributes to them. For example, the person who says life as a clarinetist is incommensurable with life as a livestock farmer might be expressing a general skepticism about judgments of intrinsic value (which obviously gives no support to Raz). Or she might be saying, in effect,

I have given some thought to this comparison, without being able to conclude with confidence either that being a clarinetist is better, or that being a farmer is better, or that they are equally valuable. I simply do not propose to worry further about the choice, now or (if I can help it) in the future.

(Note the same person, if she were sufficiently sophisticated, might add with perfect consistency that another thing she has been unable to conclude with confidence is that the choices are incommensurable. So she has not said what Raz wants.) Or she might be saying, in some contexts, "I would not presume to impose my judgment of the relative value of these alternatives on anyone else, and I will thank you not to try to impose yours on me."

In yet a different context, a person who is asked to specify the monetary value to her of one of her friendships might refuse, saying that the value of friendship simply cannot be compared with the value of money.141 I think what such a person is most likely to mean is that friendship is more valuable than any amount of money, or in other words, that the value of friendship is incomparably greater. This is

138. J. Raz, supra note 1, at 332, 341.
139. Id. at 343.
140. Id. at 336.
141. Id. at 337.
incommensurability of a sort, but it is not the sort of incommensurability Raz is asserting. Raz argues, in his later discussion of "constitutive incommensurabilities," that people often really do mean, not that friendship is incomparably more valuable than money, but that the values of friendship and money are incommensurable in Raz's sense. I find this claim quite implausible. For the moment, I shall simply hope that the reader's intuitions are with me. I shall return to constitutive incommensurabilities later on. (To avoid confusion, let me make it clear that I am not endorsing any judgments about lexical priorities among genuine values. I am merely saying it is ordinarily more plausible to interpret the claim that friendship cannot be compared with money as a claim of lexical priority than as a claim of Razean incommensurability.)

What I conclude so far is that people's claims of incommensurability or refusals to make comparisons are much harder to interpret than Raz allows.

Turning to a different argument, I suspect the crucial disagreement between Raz and me concerns the interpretation of a phenomenon about whose existence, at least under a fairly crude description, we are agreed. The phenomenon is just this: People can agonize long and hard about a career choice (for example) and end up with very little in the way of confident conclusions about what is the best decision. Raz may never argue explicitly that our frequent frustration when we attempt to make judgments of comparative value (if that is what we start out to do) shows the attempt is misguided. But I have no difficulty finding such an argument implicit in Raz—partly because it is such a strong and obvious argument. On the other hand, if the great difficulty of making confident comparative judgments argues against commensurability, there is another aspect of the phenomenon which argues in favor of it. I cannot see why, if people thought values were incommensurable, they would agonize about hard choices to the extent they often do. Their agonizing shows they believe values are commensurable. (Notice that although I am appealing to a feature of the choosing process as evidence for commensurability, I am not making the unacceptable argument that the agent will eventually make some choice, at which point she will have commensurated the values, thus demonstrating that they are commensurable. What I am saying is that unless she believes in commensurability, we cannot make sense of her agonizing before she makes her choice.)

Let me now expand. And let me consider first why, if values are commensurable, one may have such difficulty deciding whether to be (say) a clarinetist or a livestock farmer. The correct explanation is the
obvious one: even if values are commensurable, this may be a very hard choice for which we are ill-prepared. The questions involved are difficult, in a number of different ways, and neither our culture nor our teachers of philosophy have thought fit lately to offer us much help with them.

Notice first that we cannot pose the question in the form, "Which is more valuable, the life of the clarinetist or the life of the livestock farmer?" This question certainly has no determinate answer. There are many possible lives-of-a-clarinetist and many possible lives-of-a-livestock-farmer. Each class includes members that are clearly more valuable than some members of the other class. This does not amount to incommensurability; it just means we need to make our comparisons between more fully specified alternatives.

So, the first difficult part of the inquiry is trying to figure out what my life as a clarinetist and my life as a livestock farmer would be like. The life of a professional clarinetist includes many different possible activities—performing in a large orchestra, or in chamber groups, or as a soloist, teaching as a member of a music school faculty or with private students, composing or working with composers who wish to write for the clarinet, lecturing to non-professional audiences in "music appreciation" series and mingling with patrons of the arts at institutional fund raisers, perhaps becoming involved in orchestra administration, and so on. These different activities all have their own value, and one of the main things one does in trying to imagine a life for oneself as a clarinetist is to try to figure out one's chances of putting together a particular mosaic of these activities which will be feasible and will command one's interest and engagement most of the time. (This interest and engagement, I shall simply assert, is a condition not only of one's happiness, but also of the activities having any genuine intrinsic value at all.)

Interwoven with the business of figuring out just what activities of a clarinetist one might be able to be truly engaged in is the business of figuring out just how valuable these various activities really are, by systematic thought about the sources of value of musical composition, and performance, and appreciation, and of the interpersonal relationships involved in teaching or lecturing or institution building, and so on. I think such systematic thought is possible, although very few philosophers have attempted it; I think the large variety of activities and relationships we think are valuable can be seen to share certain common features, which explain and help to commensurate their value; in the end, I think the fundamental sources of value are probably reducible to one or
two. I even think that ordinary, thoughtful people look for such possibilities (in a rather disorganized way, no doubt) when choosing life projects. Or they do so at first, until they discover how complicated the questions are, and how little there is in their culture to guide them.

Of course, everything I have said about the life of the clarinetist applies equally to the life of the livestock farmer, or whatever. In sum, even if people can think hard about such questions and emerge with no comparisons they have much confidence in, we do not need to posit incommensurability as the explanation.

My sketchy description of what is involved in evaluating career choices makes it clear, incidentally, that we cannot use the "minor improvement" argument, which we used to clarify the concept of incommensurability, to prove that life as a clarinetist and as a livestock farmer are genuinely incommensurable. Imagine that I have before my mind a fairly specific life-as-a-clarinetist and a fairly specific life-as-a-farmer. I have not been able to conclude that either is better than the other. I now notice a small consideration I had overlooked, and I conclude that my life-as-a-farmer would be slightly better than I had previously thought. But I find that I am still uncertain about the choice between the clarinet and farming. What does this continued indecision prove? It proves, assuming I am rational, that my original indecision did not reflect a judgment that the life-as-a-clarinetist and the life-as-a-farmer I was considering were equal in value. It proves that, and no more. In proving I had made no judgment of equality, it does not prove I had made a judgment of nonequality (which is a necessary part of a judgment of incommensurability). The deeper reason for this, aside from the immediate logical point, is that when I said earlier I had not been able to conclude either was better than the other, that was not at all the same as saying I had been able to conclude neither was better than the other—I may well have been uncertain about everything: whether the life-as-a-clarinetist was better than the original life-as-a-farmer, or whether it was worse, or whether the two were equal. So long as the small improvement in my projected life-as-a-farmer is of the same scale as, or smaller than, the degree of uncertainty in my original judgments of value, we can perfectly well understand my continuing indecision without supposing that I regard the clarinetist's life and the farmer's life (in whichever version) as incommensurable. Lives can be hard to compare in practice without being incomparable in principle.

I have explained why, in my view, people can agonize over hard choices without achieving confident judgments of comparative value.
The question that remains is why, if values are incommensurable, people agonize at all. Raz agrees that people agonize, but it is a puzzle why, on his view, they should. He thinks it is often easy to decide that two alternatives are incommensurable. He also thinks that reason cannot guide our choice between incommensurables. But then, what am I doing when I agonize over the choice between the clarinet and livestock farming, alternatives which I ought easily to see are incommensurable? Where reason cannot guide, there is no room for practical reasoning. If there is no room for practical reasoning, should I not just flip a coin and get it over with?

One possibility, suggested by comments of Raz, is that when I agonize, I am trying to figure out how successful I will be as a clarinetist or as a farmer. I do not think this would be Raz's primary answer to the present question, but it is worth seeing that in any event it will not do. First of all, I may go on deliberating even after I have a tolerably good idea what my life as a clarinetist and my life as a farmer would be like descriptively, that is, even after I have decided how “successful” I will be in a descriptive sense. So, if I am still concerned with success, it must be with success in a normative sense. But on Raz’s view, normative comparisons in respect of success ought often to be impossible. For one thing (and let me emphasize that this point is much less important than the point that will follow), even if we consider only various lives-as-a-clarinetist, one can succeed to varying degrees in different aspects of being a clarinetist. If, as Raz says, novelists are sometimes incomparable because good novelists have different qualities, then various lives-as-a-clarinetist may be incomparable also. There may be no such thing as “degree of success as a clarinetist.” Now as to the more important point: Even if we can make sense of degrees of success as a clarinetist (and similarly of degrees of success as a farmer), how can we compare “this much success as a clarinetist” with “that much success as a farmer?” Raz does seem to assume that this sort of comparison is a possible ground for choice. But I do not see how it can be. The value of success depends on the value of what one succeeds in. If the values of the clarinet and of livestock farming are truly incommensurable, as Raz claims, then so should be the values of success as a clarinetist and of success as a farmer (nor does

142. Id. at 342.
143. Id. at 331, 334, 343.
144. Id. at 342-44.
145. Id. at 326.
146. Id. at 343-44.
147. Id. at 298.
supposing "greater" success at one, in its own terms, establish a ground for commensuration). Finally, if I am right that references to success leave the problem of incommensurability intact, then it ought often to be easy to see this. We are faced once again with the puzzle what we are doing when we go on deliberating after we see that practical reason has no more to say.

Raz suggests an answer obliquely in a passage resisting the claim that incommensurables should be regarded as "roughly equal" because there is no point in deliberating between them.148 Raz says that a choice between incommensurable careers

is the sort of decision that anyone facing it quite rightly cares a lot about. It is a choice that one ought not to be indifferent to, or unconcerned about. To be indifferent to this kind of choice is not to have a proper respect for oneself . . . . [The choice] is one of major significance for one's life.149

I am not persuaded. Of course I agree that in general the choice of a career is a choice one ought to care a lot about. And I agree that failure to care might well manifest a lack of proper respect for oneself. But Raz is not talking about the choice of a career in general. He is talking about the choice that remains after we have narrowed the field to two possibilities, and after we have decided that those two possibilities are incommensurable, that practical reason offers no further guidance. To say that we ought not to be indifferent, or that proper self-respect requires us to go on agonizing, is to say that practical reason requires us to go on deliberating beyond the point where the relevance of practical reason has been exhausted. That makes no sense. It seems to me that at this point it is perfectly appropriate to flip a coin.

I do not actually recommend flipping a coin, which would seem outrageous, because I do not think we ever get to the point Raz assumes, of being convinced we are faced with incommensurable alternatives. But if we ever did get to that point, there would be nothing better to do. Even if we went ahead and made a decision on some other ground, it would not be superior to the decision based on the coin flip. It could not be superior in outcome, since the choices are incommensurable; nor would it be superior in the deliberative process, since whatever "deliberation" we undertook could not be a species of practical reasoning. The decision process would necessarily be arbitrary. So what is the point?

148. Id. at 331-32.
149. Id. at 332.
Raz says the decision is "of major significance for one's life." What is true is that the decision will make a great deal of difference to the factual description of one's life. But if it has no consequences for one's life that practical reason can base a choice on, this appears to be a case where a large descriptive difference is not of major significance after all from an evaluative point of view. Raz might say, truly, that the choice will determine what sort of person one is for the rest of one's life. But if practical reason has nothing to say about the choice between the careers, I see no reason to suppose it has more to say about the choice between the corresponding people one might be. Once more, one might as well flip a coin.

Raz points out that if I choose to be a clarinetist, I can claim to have done so on the basis of reasons or to have chosen rationally. What he says in elaboration is that even if the clarinet and farming are incommensurable, that does not affect the fact that there are good reasons for believing the clarinet to be valuable, and for believing it to be unambiguously more valuable than some careers I might have chosen. This is true, but it is not enough to help Raz's position. In the right context, I can be said on these grounds to have had reasons for choosing the clarinet. But I cannot be said on these grounds to have had reason for choosing the clarinet as opposed to farming. That is the choice in issue.

It sounds paradoxical to claim that we should ever not take seriously a choice between careers, but that is what follows if we really believe in incommensurable values. The paradox is of Raz's making, not mine. It is Raz who says there are major choices practical reason cannot guide. I merely point out that where practical reason ends, there also is an end of morally significant practical thought. If I go on agonizing, I am waiting to be seized by unreasoned desire, or I am behaving obsessively, or some such. None of this is required by reason or by proper self-respect.

I have gotten sufficiently deep into the discussion of commensurability so that I think it is worth mentioning a possible objection to commensurability that Raz does not rely on, but that may influence some people: It may seem that if all possible lives are commensurable in value, then there is some unique best life, which we ought all to be aiming at. That conclusion is certainly unpalatable. Happily, the conclusion does not follow. One reason it does not follow is that people have different talents and inclinations to use their talents. Even if there were some unique best
"humanly possible" life, it is unlikely that the best actually feasible life for me would be the same as the best actually feasible life for you. A more important reason is this. If we remember that the goodness of lives is not, at the most fundamental level, a matter of individuals' well-being but a matter of contribution to a good "from the point of view of the universe," then it becomes highly plausible to suppose that it is intrinsically valuable that different individuals should lead a variety of different lives, even if some of these lives are less valuable than others, considered strictly in themselves. So, however highly we value philosophy (for example), we need have no fear that everyone is required to become a philosopher.

Finally (before I move on to the specific topics of constitutive incommensurabilities and loyalty), let me suggest that Raz himself may appeal implicitly to commensurability in reaching his political conclusions. Raz thinks government is justified in taxing the rich in order to increase the opportunities of the poor.151 (And so do I.) This seems to depend on an assumption that the difference the extra money makes to the value of the lives of the rich is less than the difference it will make to the value of the lives of the poor. How can we say that, if most career choices (and other activities also) are incommensurable? For reasons I have already discussed, it is not enough just to say the money will make less difference to the overall success of the lives of the rich. The rich and the poor may choose different kinds of lives, and lives built on incommensurable pursuits cannot be commensurated by reference to degrees of success. (Nor, for similar reasons, can different kinds of lives be commensurated by degrees of autonomy. Just as the value of success depends on what one is successful at, so does the value of autonomy depend on what one does with it.)152 In order to claim that the redistribution increases overall well-being, it is necessary (at least) to make some assumptions about how the distribution over kinds of life of the choices made by the poor compares to the distribution over kinds of life of the choices made by the rich. I should say we need (at least) the assumption that every incommensurable kind of life is chosen by at least as many people post-redistribution as pre-redistribution. There is room for argument about whether this assumption is plausible; but Raz mentions no such assumption at all. The ease with which Raz assumes that redistribution will increase overall well-being suggests to me that he has forgotten about incommensurability, and is rather assuming that "standardly

151. Id. at 415-17.
152. See infra text accompanying notes 203-04.
successful" lives devoted to various careers and activities really are comparable in value, and are approximately equal in value at that.\textsuperscript{153}

There is a similar problem in Raz's very interesting discussion of toleration.\textsuperscript{154} According to Raz, when we exercise toleration, we resist the expression in behavior of an \textit{intrinsically valuable} intolerant attitude. The intolerant attitude is valuable because it is part of some genuine virtue we possess. I assume that the expression in behavior of this intolerant attitude would also, considered in itself, be intrinsically valuable. (Raz certainly assumes in some cases that the expression of an intrinsically valuable attitude is intrinsically valuable; and he suggests no reason why this should not be taken as a general principle.) Yet we restrain the behavior, lest it interfere with valuable activity by others with different virtues from our own. Perhaps this is a correct description of toleration, but even assuming that it is, we can still ask the question why should we be tolerant? I suspect Raz implicitly assumes that we lose less value by forgoing our expressive behavior than we gain by allowing others' activities to proceed. But I do not see how he can assume this without assuming values are commensurable.

Raz might respond that our reason for being tolerant is not that we thereby increase overall value. Rather, our reason is just that toleration is a virtue, with a value of its own. And toleration requires certain sacrifices of us, which we make without looking to an overall value, just as friendship does. (Indeed, toleration might be thought of as an aspect of civic friendship.) There are two consequences of taking this line. First, we are left without an \textit{argument} for toleration; we are left relying on the bare intuition that toleration is a virtue. For myself, I have no objection in principle to such reliance on intuition (though I do not share this intuition, in just Raz's version of it). But I suspect many believers in toleration would want to have an argument. Second, if all Raz has to say in favor of toleration is that it is a virtue like friendship, then he clearly has

\textsuperscript{153} A further complication, treated very briefly: There is a passage in which Raz suggests that we might make interpersonal comparisons of success, even among people who choose very different lives, by asking how well each did relative to the value of the \textit{opportunities} available to her. J. Raz, \textit{supra} note 1, at 299. Now, I doubt that we can answer in any useful way the question how well each individual did relative to the opportunities available to her, since she presumably had some choice among incommensurable activities, and we have no grounds for assuming it is more valuable to do slightly better at one of two incommensurable activities than to do slightly worse at the other. But notice that even if we could make the interpersonal comparison of success that Raz suggests, it would not help with the argument for redistribution. If success is what matters, and if one's success is a matter of how well one does relative to one's opportunities, then the poor are in no way disadvantaged by having lesser opportunities, and we have no argument for trying to increase them.

\textsuperscript{154} \textit{Id.} at 401-07.
no ground for saying that everyone ought to be tolerant. But I think he wants to say that.

Now back to the specific topics I mentioned earlier, of which the first is constitutive incommensurability. It is simplest to state Raz's claims in the context of a specific example, so let us take friendship. Raz argues that it is constitutive of a capacity for friendship, as we understand it in our society, that (a) one should view the values of friendship and at least some other genuine values as incommensurable, and (b) one should be unwilling, if one actually is a friend, to give up the values of friendship for any of those other incommensurable values. If Raz were right about this, then constitutive incommensurabilities would be substantial evidence that people actually believe in the incommensurability of some values, which would in turn be some evidence for the truth of incommensurability. But it seems to me that Raz misdescribes our views about friendship and similar relationships.

Some pages back, I mentioned Raz's case of a person who refuses to specify the monetary value to her of her friendships. I suggested that what this person probably means by her refusal, and what she would most likely say in explanation, is that her friendship is worth more than any amount of money. Raz notices this possible objection, and he attempts to deal with it in a discussion of parenthood.

Most of us would agree that parents should be unwilling to sell their children, and indeed that it is constitutive of being a parent as we understand it (not in the biological sense but in the "real" sense) that one should be unwilling to sell one's children. Now, by itself, a refusal to sell one's children could be explained by the simple belief that having children is worth more than any amount of money. There is no incommensurability in that belief. But Raz thinks that for many people (and, by implication, for the right-minded among us, at least in the context of our society), buying children is as bad as selling them. "[E]xchanging [children] for money, whether buying or selling, is inconsistent with a proper appreciation of the value ofparenthood." This revulsion at buying as
well as selling shows that what moves parents is not a belief that having children is worth more than money, but rather a belief that the values of having children and of money are incommensurable.

But do we really believe that a willingness to buy a child is inconsistent with a proper appreciation of the values of parenthood? Consider one of those unfortunate couples who are incurably infertile, who are turned down on some ground or other by various adoption agencies to whom they apply for a child to adopt, and who finally in desperation buy, or attempt to buy, a child in a transaction which they may or may not know is illegal? Does this couple demonstrate by their behavior a misapprehension of the value of parenthood? I hardly think so.

To be sure, if they think they can become parents in the true sense just by paying out money and taking physical custody of the child, then they are grievously in error. But there is no reason to assume they think that. Also, we may disapprove of this couple's behavior on the ground that they contribute to the continuation of an illegal traffic in children which is (we may assume) most undesirable on the whole. Even if the couple's behavior merits disapproval on this ground, it still does not show they lack a true appreciation of the nature or value of parenthood. Many people are not attuned to this sort of argument from large scale consequences, and the couple may have persuaded themselves that the particular child they are buying will be better off with them than in the hands of anyone who is willing to sell it. (About this last point, they may well be right.) Possibly we disapprove of the couple's behavior because we suspect that if a number of adoption agencies denied them a child there is some good reason they ought not to have one. But of course, official agencies do make mistakes. And in any event, misestimating one's own capacity for parenthood is not the same thing as failing to appreciate the true value of parenthood.

There are further reasons why we regard the whole idea of buying children as unsavory. Partly, we think it unfortunate that anything so significant to people's lives as the opportunity for parenthood ought to depend on having money. More important is a general negative symbolic significance of money, and therefore of monetary exchanges, in moral contexts. I shall come back to this symbolic significance, which Raz notices but which I think he misdescribes, in a moment. But for now, I think the example of the infertile couple suffices to show that buying a child does not necessarily manifest a failure to appreciate the nature or value of parenthood. I am left still believing it is closer to the truth to
say we regard the value of parenthood as incomparably greater than the value of money, than to say we regard these values as incommensurable.

Most of Raz's discussion of constitutive incommensurabilities involves not parenthood but friendship. Raz thinks it is constitutive of a capacity for friendship that one regards friendship and money as incommensurable, and that one who has chosen to be a friend would not exchange the values of friendship for money. Raz is led by his analysis to say some things that strike me as quite peculiar. For example, "Only those who hold the view that friendship is neither better nor worse than money, but is simply not comparable to money or other commodities are capable of having friends." This entails that someone who really does believe friendship is more valuable than any amount of money is incapable of having friends. One could argue in defense of this that one shows his friendship even more by preferring friendship to other values which are not lesser (because they are incomparable) than by preferring friendship to other values which are definitely less. Still, it strikes me as very odd to say that if I think friendship is more valuable than any amount of money, then I cannot be a true friend.

Later on Raz says most of us on reflection do not believe that people who forgo the possibility of friendship for money act wrongly and against reason. Rather, we "regard such people as one regards people who neglected their musical abilities and lost their sensitivity to music in order to be better able to pursue some other goals." On the contrary, I think most of us do regard people who forgo friendship for money as acting wrongly and unreasonably. (And we would feel the same way about someone who gave up her musical sensitivity just to pursue money.) There is an even stranger consequence. Raz's claim about incommensurability treats friendship and money symmetrically. That suggests that the person who opts for friendship is making a choice which has a significant cost: She is giving up the opportunity for a deep and meaningful relationship with money.

---

159. *Id.* at 352.

160. Perhaps I should admit that if I started from the premise of incommensurability, then I might be moved to endorse the "very odd" claim in the text — influenced by the general view that one's engagement in a project cannot be valuable unless one understands, at least to some required degree, the nature of the value of that project. But still, the counter-intuitiveness of the very odd claim counts against a belief in incommensurability. We need particularly good independent reasons to believe in incommensurability if it has consequences like this.

161. *Id.* at 352-53.
A major problem with Raz’s discussion is just that it is conducted in terms of exchanges of other values for money. Money is sometimes valued for itself (mistakenly, but people do it), sometimes valued as a means to other goals of slight (or zero, or negative) value, and sometimes valued as a means to other goals of substantial positive value. So, when we ask about the reasonableness of exchanging other values for money, it is important to be clear just how the money is being valued. It seems to me that Raz is not always clear about this.

Most commonly, when we talk in the course of a moral discussion about doing things or exchanging things for money, we are implicitly assuming that the motivation is base. It is pure avarice, or inappropriately materialistic desires, or a desire for the crude social prestige money may bring, or whatever. Money has become a symbol, in such discussion, for a wide range of bad motives. This is why most of us could easily find ourselves saying in some context that friendship is more valuable than any amount of money. Money, as it is normally considered in this kind of context, has no value at all, or negative value.

Raz notices the special symbolic significance of money, but he suggests (for the most part) that this significance is internal to those relationships which are partly constituted by a belief in the incommensurability of the values they create and the value of money. Thus, “what has symbolic significance is the very judgment that companionship is incommensurable with money.” But the symbolic significance of money for moral discussion is by no means limited to specific contexts involving friendship, or marriage, or other relationships that might be thought to involve constitutive incommensurabilities. It is quite general. So, our distaste for imagined monetary transactions is not a matter of constitutive incommensurability, but of the base motivation which reference to money, in this sort of discussion, usually connotes.

Suppose we are careful to imagine a case where a friend or a spouse faces a choice between the value of companionship and a sum of money regarded as a means to some other genuinely valuable goal, say perhaps an excellent violin for a devoted chamber music player? Is it constitutive of being a friend or a spouse that one would not consider separating oneself from one’s friend or spouse for a month in order to secure the money for the violin? It does not seem so to me.

Even Raz seems to waver. Raz explicitly allows that someone might leave her spouse for a month in order to earn money, but he claims the

162. Id. at 350 (emphasis omitted).
same person would not leave her spouse for the same month in return for a gift from a stranger of a larger amount of money. Now, if we assume that the money is desired in both cases for some further valuable end, this behavior seems prima facie irrational. The spouse who would go away to work for a month to make money for a violin (say), ought also to be willing just to go away for a month if that would equally secure the money and the violin. If the unwillingness just to go away is not irrational (and I agree it might not be), it will be because the stranger offering the money has contrived to force a particular symbolic significance on this particular exchange, that makes it a denial of the value of companionship. It is not because an exchange of companionship for a sum of money (valued as a means to other valuable ends, remember) necessarily has an objectionable symbolic significance. (It might seem that every conditional gift from a stranger would have such a significance, but I do not think even that is true. Without describing cases in great detail, imagine that the source of the gift is a rich and eccentric old music lover who imposes the condition, misguidedly no doubt, as a test of the recipient's devotion to her music.)

In allowing the spouse to go away to work for a month to earn money, Raz undermines much of what he appears to say about constitutive incommensurability. It is easy to get the impression that the constitutive incommensurability of friendship and money forbids a friend from ever sacrificing companionship for money. That is the most natural interpretation of the claim that "only those who would not even consider exchanges of money for friendship are capable of having friends" in its immediate context. But if it is acceptable for someone to leave her spouse to earn money, then it seems all the constitutive incommensurability really forbids is some incompletely specified class of monetary exchanges with strangers. So limited, constitutive incommensurabilities may not lose all their interest. But the claim that we believe in such "constitutive incommensurabilities," even if it is true, becomes much less weighty evidence for the claim that we believe in truly incommensurable values in general. The narrower the class of exchanges that is prohibited, the less the prohibition seems to reflect any deep view about the nature of the underlying values.

I turn now to the final topic for this section, Raz's discussion of loyalty and commitment. Raz observes that his treatment of constitutive incommensurability gives rise to a puzzle: If the values of friendship and

---

163. Id. at 348-49.
164. Id. at 352.
money really are incommensurable, why do we regard it as wrong to give up a friend for money? Why can we not switch our allegiance freely between incommensurable values?

Let me emphasize that although I do not believe in incommensurability, and although I shall also disagree with Raz’s account of loyalty, I still think there is a genuine puzzle of the sort Raz points to for any consequentialist who has a plausibly complex view of what activities and relationships are intrinsically valuable. When we take up some valuable activity or enter some valuable relationship, we do seem to be making a commitment in some sense. It is harder to back away honorably from such an activity or relationship than it is to avoid it in the first place. (Incidentally, the fact that activities like playing the violin, and not just personal relationships, are “sticky” in this sense makes it clear that the stickiness involves something more than just not disappointing other people’s legitimate expectations.) Raz does well to raise the puzzle, and I think his remarks suggest the outline of a correct solution to it. But I think Raz goes further than necessary both in stating the puzzle (which involves for him the existence of constitutive incommensurability) and in answering it.

Raz’s answer to the puzzle is duties of loyalty, which apply only to those who have made a commitment to one or another of two incommensurable values. But why do we have such duties? Raz’s immediate answer is that the duty of loyalty to friends (for example) is created by the conventions of friendship. “All social forms involve ways of being true to the project or to the relationship which they define.” But, unless we are thoroughgoing conventionalists, the social forms, even if they include among their terms requirements of loyalty, cannot create duties of loyalty all by themselves.

What about our voluntarily undertaking projects or relationships whose terms include requirements of loyalty? Does not that create duties? As Raz says, “Being true to pursuits and relations is being engaged in them according to their terms.” Now, we can view it as a logical truth that being true to a pursuit is being engaged in it according to its terms. But it is far from a logical truth that we ought always to be true to our pursuits as conventionally constituted. Imagine that I live in a society where one of the conventions of friendship is that I should always take my friend’s part in any controversy with a foreigner, even to

165. Id. at 356.
166. Id. at 354.
167. Id. at 355.
the point of lying or using violence on my friend's behalf. I take it we would all agree that if this is the convention of friendship in my society, then there are cases where I should violate the convention, even if my behavior is conventionally regarded as disloyalty.

Raz might respond that this is a case where I ought indeed to be disloyal, but where one of the effects of my disloyalty is that I am (or I recognize that I am) incapable of friendship. But this is surely too strong. Friendship is not a purely conventional notion, and to be incapable of friendship according to the conventions of my society is not to be incapable of friendship *simpliciter*. What is true (as Raz once pointed out to me in conversation) is that I will experience difficulty forming genuine friendships with other members of my society but outside the accepted conventions of my society. We rely on conventions to shape our projects and relationships, and we cannot create new conventions at will. On the other hand, conventions do begin somehow, and grow and change; and friendship, as I say, is not a purely conventional notion; so we cannot exclude out of hand the possibility that I will be able to form friendships, whose terms will, as a practical matter, be close to the conventional terms of friendship in my society, but which need not be identical.

In sum, we cannot say that any behavior my society brands as disloyalty in a friend is wrong, nor can we say that it incapacitates me for friendship. We could still assert that every conventional disloyalty violates a duty (recognizing that duties may be overridden), but if the rest of what I have said is correct, I hardly see the point in doing so.

Of course Raz never says that we should commit violence on behalf of our friends if our society requires it, and perhaps he would admit that *this* kind of "disloyalty" does not even incapacitate one for friendship. Can he retreat to the claim that we are bound by conventional duties of loyalty where, and only where, those take the form of requiring us to prefer one of two incommensurable values to another if we have made a commitment to the former? This claim is more plausible, but I still think it attributes too much force both to the conventions and to commitments. Conventions by themselves settle no question of moral significance. As to commitments, I shall discuss later on, in the section on autonomy, both Raz's view that commitments are in themselves reasons for action and my objections to that view.\textsuperscript{169}

\begin{itemize}
\item \textsuperscript{168} Id. at 355-56.
\item \textsuperscript{169} See infra text accompanying notes 185-92.
\end{itemize}
I began this discussion of loyalty and commitment by saying I
thought there was a genuine puzzle about the "stickiness" of valuable
projects, to which Raz suggested the outlines of a proper solution. So,
after all this disagreeing with Raz, it is time to say just how I agree with
him.

I think Raz is right (in general, not just in this section of his book)
that actions may have intrinsic value or disvalue, which must be taken
into account in deciding how to act. He is right in thinking that the
intrinsic value of actions ("some actions," for Raz) "depends on the spirit
in which, or on the reasons for which, they are undertaken."170 More
specifically, actions have their intrinsic value or disvalue largely in virtue
of the moral attitudes which they manifest. What moral attitudes an
action normally manifests depends to a considerable extent on the con-
ventions of the society in which the action is done. And that is one rea-
son why Raz is quite properly interested in conventions. But this is the
point at which Raz also goes too far. Neither our actions nor our atti-
tudes are ever entirely constituted by social conventions. There is always
the possibility that in relying on convention to interpret an action we will
misinterpret it, and so misevaluate it.

If the value of our actions is not definitively determined by conven-
tional interpretation, then neither can the stickiness of projects (includ-
ing relationships) be entirely a matter of convention. The explanation I
would give for the stickiness of projects goes approximately as follows: It
is harder to back away from a project one has made a "commitment" to
than not to take up the project in the first place because the process of
making the commitment has deepened one's appreciation of the value of
the project. (This would not be true if one could make a commitment to,
say, playing the violin just by announcing it as one's project; but one
cannot. One cannot make a real commitment to playing the violin with-
out making some attempt to learn to play it. And probably immersing
oneself in the culture of the violin in other ways as well.) The more fully
one appreciates a particular value, the fewer are the circumstances in
which one will feel justified, or will be justified, in choosing against that
value to pursue another. And therefore, the more one has committed
oneself to a project, the fewer are the circumstances in which one may
choose against it. This line of argument, which obviously requires elabo-
ration, promises to explain the stickiness of projects without suggesting
in any way that commitment gives rise to new "duties" (although it does
change what one ought to do), and without suggesting any reason to

170. Id. at 363 (emphasis added).
think that choosing against a value one is committed to incapacitates one from future pursuit of that value even if one's choice was justified.

D. AUTONOMY

The last two chapters of Raz's book are entitled "Autonomy and Pluralism" and "Freedom and Autonomy." Claims about the value of autonomy are the principal link between Raz's moral theory and his political conclusions. In this section I shall not directly address Raz's political conclusions, though I shall comment on a couple of points that come up along the way. My main concern is with the question whether, and in what sense, autonomy is valuable.

First off, we should be clear what Raz means by autonomy. Autonomy is a matter of making significant choices between comprehensive goals, choices that shape one's life. As Raz says, "The autonomous person is a (part) author of his own life." This notion of autonomy Raz sometimes refers to as "personal autonomy," to distinguish it from moral autonomy. Moral autonomy, an idea familiar since Kant, is something like acting under laws one has legislated for oneself. But moral autonomy does not necessarily entail personal autonomy. It is possible, for example, to be morally autonomous without having any real choice about how one's life is to go if it is to be a moral life at all. The personally autonomous person does have such choices. He faces a range of valuable options. "Moral autonomy both in the Kantian and in other versions is a doctrine about the nature of morality. Personal autonomy is no more than one specific moral ideal which, if valid, is one element in a moral doctrine." Hereafter, when I speak of "autonomy" I shall mean personal autonomy.

Personal autonomy, as defined by Raz, is also different from merely having goals that one willingly embraces. It is perfectly possible that some agent should be coerced initially into undertaking certain projects, but should eventually come to be fully absorbed in those projects, to willingly embrace them as her own. Even if this turns out to be the case, the projects have not been autonomously chosen, and the person who acquires all of her major projects in this way is not autonomous. She is not the author of her own life. (Raz believes, incidentally, that an agent's

171. J. Raz, supra note 1, at 369-99.
172. Id. at 400-29.
173. Id. at 369.
174. Id. at 370 n.2.
175. Id. at 369.
willingly embracing her projects is a necessary condition for those projects to be valuable. And I agree. But willingly embracing is not a sufficient condition for the projects to be autonomously chosen.)\textsuperscript{176}

Raz spends some time spelling out the conditions for autonomy. First, in order to be autonomous, a person must have certain “inner capacities required for the conduct of an autonomous life”—cognitive and affective capacities, health and physical abilities, and certain character traits. In addition, she must have an adequate range of valuable options from which to choose. Finally, she must be free from coercion or manipulation by others.\textsuperscript{177} But being autonomous involves more than satisfying the conditions or having the capacity for autonomy. Autonomy is an achievement, not just a capacity. It is possible to have the capacity for autonomy and yet not to have an autonomous life. One can drift through one’s life, never making genuine commitments to any projects at all, perhaps even deceiving oneself into thinking one has made such commitments when in fact one has not.\textsuperscript{178}

I fully agree with Raz that to the extent autonomy is important, what is important is achieved autonomy, not a mere capacity. This, incidentally, is Raz’s most powerful argument against people who want to base moral or political theory on \textit{rights} to autonomy. Talk of rights belongs with the idea that what matters is the capacity for autonomy, not the achievement.\textsuperscript{179} But as Raz points out, we value the capacity only because we value the achievement for which it is necessary.\textsuperscript{180}

Enough of preliminaries. Now to the question, is autonomy valuable? Is it valuable in such a way that it ought to be a general goal of political action?

Let me summarize my view of Raz’s discussion, before I launch into details. Raz says a great deal that is true and interesting in connection with the value of autonomy (as he does with every topic he discusses). Nonetheless, I think there is a basic defect of the discussion. Raz’s principal argument for the value of autonomy turns out to be an argument for its being instrumentally valuable, or at most for its being valuable as a constituent of well-being \textit{under certain conditions}. Indeed, Raz is quite explicit, on occasion, that autonomy is not valuable under all conditions. “I do agree . . . that not everyone has an interest in personal autonomy.

\begin{itemize}
\item \textsuperscript{176} \textit{Id.}
\item \textsuperscript{177} \textit{Id. at} 407-08, 373-78.
\item \textsuperscript{178} \textit{Id. at} 369-73, 381-83.
\item \textsuperscript{179} \textit{Id. at} 204-05.
\item \textsuperscript{180} \textit{Id. at} 372.
\end{itemize}
It is a cultural value, i.e. of value to people living in certain societies only."\textsuperscript{181} "Autonomy is . . . inconsistent with [some] valuable lives."\textsuperscript{182} However, having established a conditional claim for the value of autonomy, Raz then seems to go on and argue for further propositions, appealing to the value of autonomy as if it were unconditional.

First, let us consider the arguments for the value of autonomy. The principal argument is just that "[f]or those who live in an autonomy-supporting environment [such as ours] there is no choice but to be autonomous: there is no other way to prosper in such a society."\textsuperscript{183} In our society, we have abandoned the practice of arranged marriages; we have abandoned the assumption that a child will follow its parent's occupation in life; and so on. Our social forms presuppose that people will make many of the important choices about the shape of their lives. The person who waits for society to shape a life for her will have very little of a life at all.

So far, the argument for autonomy may sound purely instrumental, but, as I said, it is more than that. Raz argues persuasively that the social forms themselves are affected by the emphasis on personal choice. Marriage means something different in a society where people are expected to choose their marriage partners than in a society where marriages are arranged. People expect different satisfactions from their jobs in a society where they choose them for themselves. And so on. So it is not just that the nonautonomous individual in a society like ours will never get started on life's most important projects. That problem might be solved by a little well-directed coercion, at least in theory. The deeper problem is that the nonautonomous individual is totally unfit to participate in many of the social forms of our society. The forms have changed in response to the fact that they are now for the most part objects of deliberate choice; and they have changed in such a way that only people capable of making serious choices for themselves can take part in them at all.\textsuperscript{184} This is why autonomy is not merely of instrumental value. It is a constituent of well-being.

This argument is perfectly sound. But what it establishes is that autonomy is a constituent of well-being \textit{in a society like ours}. Before I go on to explain how I think Raz misuses this conclusion, treating it as if it established the unconditional value of autonomy, let me pause over one

\textsuperscript{181} \textit{Id.} at 189 n.1.
\textsuperscript{182} \textit{Id.} at 395.
\textsuperscript{183} \textit{Id.} at 391.
\textsuperscript{184} \textit{Id.} at 390-95.
other argument for autonomy someone might find in Raz. I am not cer-
tain whether Raz himself intends it as an argument for the value of
autonomy; but whether he does or not, it raises issues of substantial
independent interest.

A view Raz holds is that commitments constitute, in themselves,
reasons for action. Consider an example. If I am not currently a violin-
ist (not even an amateur), I have some reason to take up the violin, just
because playing the violin is a valuable activity. Even so, if I never take
up the violin, that is not a failure on my part, nor does it damage my
well-being. But if I do take up the violin, and make a genuine commit-
ment to learning to play at a decent amateur level, then I have a new
reason for working on the violin over and above the reason I had before I
made the commitment. I also have new possibilities in my life for success
or failure: I can now succeed by living up to, or fail by not living up to,
my commitment.185

Now, as I said, it is not clear just what connection Raz sees between
this doctrine of commitments-as-reasons and the value of autonomy. But
Raz's remarks could certainly suggest to someone the following argu-
ment: Because commitments, if viewed as reasons, give us new opportu-
nities for success and failure, it is valuable to be able to make
commitments. It is not possible to make commitments in the relevant
sense unless one is autonomous. Therefore autonomy is valuable.

For myself, I am not at all persuaded that commitments are new
reasons for action. I admitted in the previous section that there is the
phenomenon I refer to as the "stickiness" of commitments. But if the
explanation I sketched (very superficially) for that stickiness is correct,
then making a commitment will normally change one's reasons for
action, without the commitment itself being a new reason for action.
(Notice the parallel here with the disagreement between Raz and me
about how authoritative utterances solve coordination problems.) Raz
tries to justify the claim that commitments are reasons for action by an
analogy to promising.186 Once again, my view would be that promises
(usually) change one's reasons for action—and indeed are made with the
purpose of changing one's reasons for action—but that they accomplish
this without being reasons for action in themselves. (Raz's analogy to
promising is problematic for another reason as well. I suspect most peo-
ples, including many who ordinarily incline more to Raz's view on all of
this than to mine, would think the validity of a promise as a reason

185. Id. at 386-89.
186. Id. at 387-88.
depends on the fact that interests of other people than the promisor are involved. But a commitment to playing the violin may involve no one but the agent.)

Raz might conceivably argue that commitments are reasons because it is desirable that commitments should be reasons: the notion of commitments as reasons is what grounds the idea that making commitments creates new opportunities for success or failure in our lives. The basic form of argument here—that commitments are reasons because it is desirable that they should be reasons—seems to me no more valid than it did when it was used to establish the validity of consent to authority. In addition, I doubt the premise that it is desirable that commitments be reasons. I cannot see that we add much to the goodness of human life by regarding the opportunity to succeed or fail in living up to our commitments as something additional to the opportunities we have simply to succeed or fail in valuable projects. (Most of us do much less well than we ought just at exploiting those basic underlying opportunities.)

Raz might retort that, conceptually, there is no such thing as “success” or “failure” in the absence of commitment. To this I have two answers. First, I think this is just not true, in the sense of commitment relevant to the present discussion. One can, for example, succeed or fail in a project that one has been coerced into and has then willingly embraced. But there is no “commitment” here in the presently relevant sense of commitment, which is “the commitment of an autonomous person.” (Remember it was an argument for the value of autonomy that got us on to this topic.) In our society, of course, one who waits to be coerced into valuable projects may wait in vain. So it may be necessary to make commitments as the only way of getting involved in projects at which one can then succeed or fail. But this does not make commitment conceptually necessary to success or failure; nor does it give us any reason to value the opportunity to succeed or fail in the commitment as opposed to the mere opportunity to succeed or fail in the project.

Second, if anyone still thinks success and failure do depend conceptually on commitment in Raz’s sense, then I am quite prepared to say that the value of a life is not to be measured by “successes” and “failures.” Rather, it is to be measured by the value of the willingly embraced and effectively pursued activities and relationships it contains. Whether these valuable activities and relationships count specifically as

187. See supra text accompanying notes 80-95.
188. Cf. J. RAZ, supra note 1, at 386-87.
“successes” is irrelevant. 189

Before leaving the topic of commitments-as-reasons, let me bring forward one last idea, without developing it. Raz believes in the incommensurability of values. I do not remember that he ever says that because he believes in the incommensurability of values, he must believe in commitments as reasons. But it may be true. (Raz may approach this idea when he says that “The typical role of our decisions and choices... is to settle what was, prior to our commitment, unsettled. It makes the pursuit of the embraced option the right pursuit for us.”) 190 The possibility I want to draw attention to is that Raz needs the doctrine of commitments-as-reasons in order to explain why, having made a decision between incommensurable values, one then follows through on it. If one believes values are commensurable, as I do, then after one has thought about a problem, one can say to oneself, “I have done the best I can to identify the right course. More thought will not produce a better result. So let’s get on with it.” One treats one’s decision as authoritative for oneself at later times (and this can involve merely my indicator-rule conception of authority). The believer in incommensurability cannot say quite the same thing. He was never trying to identify “the right course.” He can tell himself that more thought will not produce a better result; but if Raz’s picture is correct, he has already thought well beyond the point where the object was to produce a better result. 191 How then can he resist rethinking the whole matter continuously unless he can make a commitment which functions as a new reason? 192

189. It might be objected that this view leaves us no way to account specifically for “failures,” which now matter only indirectly, because of the pain of frustration or the opportunity cost of the personal resources fruitlessly invested. This objection does not move me. I see no reason to think that mere failures at valuable projects should count directly as positive demerits in a life. What we might refer to as “dishonorable abandonments” of projects do count as positive demerits. But not every failure, in the ordinary sense of “failure,” is a dishonorable abandonment. And if it be said that I have now conceded the significance of at least one category, namely “dishonorable abandonment,” which we cannot understand except by reference to the importance of commitments, I answer that that is still not really the case. The significant category here which brings demerit into a life is something like “disrespect for a genuine value.” Insofar as commitments, as I see them, create new reasons for action (without constituting new reasons), they create new particular modes of disrespecting genuine values; and these new modes that depend on commitments we can collect under the label “dishonorable abandonment.” But the really significant category is still “disrespecting genuine values.” There are ways of disrespecting values that one has never made any commitment to. So, commitment is not conceptually tied to the category that is morally fundamental.

190. J. RAZ, supra note 1, at 389.

191. See supra text accompanying notes 142-50.

192. There is one other possible argument for the unconditional value of autonomy, which I am certain Raz does not intend as an independent argument, but which some reader might think significant. At one point Raz states that “To be autonomous... a person must have options which enable him to sustain throughout his life activities which, taken together, exercise all the capacities human
So much for arguments about the value of autonomy. Raz has given a sound argument for the conditional value of autonomy in a society like ours, and no more. (Remember that when he is being careful, he does not even claim more.)\(^{193}\) The next question is: Why do I say that Raz seems to make arguments which presuppose a stronger claim than this about the value of autonomy?

Probably the most salient example of such an argument is the following: Raz claims that society has a duty to guarantee for people the conditions of autonomy. He claims also that the conditions of autonomy include an adequate range of valuable comprehensive goals to choose from. Therefore, he concludes, society must guarantee an adequate range of valuable goals.\(^{194}\)

This would make perfect sense if the value of autonomy were unconditional. But it sounds distinctly odd if the value of autonomy is conditional. If autonomy is valuable only because one needs it to participate in the options one's society offers, then it is hard to see how society could fail in any autonomy-based obligation by providing too few options. Presumably the autonomy provided by the existence of the existing options is what one needs to take advantage of the existing options (at least insofar as we are focusing on that particular aspect of autonomy which is having a range of options). But if whatever range of options society offers is enough, there is no way it could offer too few. And if that is so, there seems to be no point to saying society has an obligation to offer an adequate range of options.\(^{195}\)

Closely related to Raz's claim about an adequate range of options is a claim about moral pluralism. Raz argues that in order to be autonomous one must face a range of choices which differ in morally significant

---

\(^{193}\) See supra text accompanying notes 181-83.

\(^{194}\) J. Raz, supra note 1, at 417-18.

\(^{195}\) Raz points out that a society should not abolish currently available life options too quickly, since that prevents some people from having the lives they have already chosen. Id. at 411. That establishes a sort of connection between valuing autonomy and providing a range of options. But this is a much weaker claim than Raz normally seems to be arguing for. He normally seems to be arguing that society owes everyone, including those just entering on adult life who have not yet made significant commitments, a wide range of options from which to choose.
ways, and from this he deduces that valuing autonomy commits one to value pluralism.\textsuperscript{196} The argument makes sense (though I do not mean to concede it is clearly sound) if autonomy is unconditionally valuable. But again, it makes very little sense if autonomy is valuable merely because it is necessary to prosper in a particular society. If that is why autonomy is valuable, it would seem that whether valuing autonomy commits one to value pluralism depends on whether the society which conditions the value of autonomy recognizes plural values. And if the connection between autonomy and value pluralism depends on independent facts about society, then even if the society does recognize plural values, so that the required autonomy involves choosing among plural values, it would be odd to say that it was our valuing autonomy that grounded the belief in value pluralism.

I turn now to a slightly different way in which Raz over-relies on the claim that autonomy is valuable. Raz wants to defend something like John Stuart Mill's harm principle as a limitation on government coercion. The harm principle does not mean, for Raz, that government has no business enforcing morality. In a sense, government's entire business is enforcing morality. Even so, Raz wants to argue that government should not use coercion or manipulation to stop people from engaging in worthless, or even evil, behavior, provided the behavior harms no one.\textsuperscript{197}

Raz's principal argument for this claim is that coercion is such a clumsy tool. "[T]here is no practical way of ensuring that the coercion will restrict the victims' choice of repugnant options but will not interfere with their other choices."\textsuperscript{198} This argument is not obviously successful even if coercion takes the form of imprisonment (Raz's strongest case). After all, to the extent the threat of imprisonment is successful, there will be no imprisonment and no interference with other choices at all. When coercion takes less invasive forms, or when it addresses itself directly only to the providers of opportunities for repugnant options, the argument is weaker still. And it seems to me that the argument is attenuated to the vanishing point when we shift the focus from coercion to manipulation. (Raz claims that "Manipulating people . . . interferes with their autonomy, and does so in much the same way and to the same degree, as coercing them."\textsuperscript{199}) Suppose, for example, we think that extensive playing of video games is a worthless activity (which I am not quite inclined

\begin{flushleft}
\textsuperscript{196} Id. at 398-99. \\
\textsuperscript{197} Id. at 403-04, 418-19. \\
\textsuperscript{198} Id. at 419. \\
\textsuperscript{199} Id. at 420.
\end{flushleft}
to believe) or that it is much less valuable than other things people might be doing with their time (which I am inclined to believe). If we can discourage the playing of video games by the same sort of advertising that leads people to think smoking cigarettes will make them more up-to-date or more sexually attractive, then we have a technique for discouraging video gaming which, even though it is crassly manipulative, interferes with no other activity at all.

I am not saying there is nothing wrong with manipulation. In some ways manipulation is more objectionable than coercion just because the person being influenced is more likely to be unaware what is being done to him. Also, manipulation normally exploits some undesirable tendency in people (such as the desire to have the latest fashionable car or to look like the latest fashionable movie star), and every appeal to these tendencies may reinforce them.

Still, if there is to be the very strong presumption that Raz asserts against manipulating people (or even coercing them) to avoid worthless or evil (but harmless) activities, I do not think it can be based primarily on the "clumsy tool" argument. It must rely heavily on the likelihood that government will use coercion and manipulation when there are better methods such as education available, or that government will use coercion and manipulation to the wrong ends.  

There is a more general theme interwoven with Raz's clumsy tool argument, and it is actually this more general theme that is my final example of Raz's over-reliance on the value of autonomy. The theme is that autonomy is autonomy even when it is exercised in bad or worthless choices, and that every invasion of autonomy is presumptively to be avoided.

A moral theory which values autonomy highly can justify restricting the autonomy of one person for the sake of the greater autonomy of others or even of that person himself in the future. . . . But it will not tolerate coercion for other reasons. The availability of repugnant options, and even their free pursuit by individuals, does not detract from their autonomy.  

Elsewhere, "A person is autonomous even if he chooses the bad."  

But is it true that a person is autonomous even if he chooses the bad? Is it true that interfering with bad choices interferes with autonomy? For most people, of course, these are easy and straightforward

---

200. Cf. id. at 427-28 (discussing dangers inherent in the concentration of power in few hands).
201. Id. at 419.
202. Id. at 411.
claims. For Raz, however, they are not. (And it is a good feature of Raz's position that they are not.)

Raz says repeatedly (and correctly) that "[a]utonomy is valuable only if exercised in pursuit of the good." Or, "[a]utonomous life is valuable only if it is spent in the pursuit of acceptable and valuable projects and relationships." Now, there is no inconsistency in saying that autonomy is autonomy even when it chooses the bad, and saying also that autonomy is valuable only when it chooses the good. But there is an inconsistency in asserting both of these propositions and in supposing also (as Raz does) that autonomy, tout court, is valuable. These three propositions taken together entail that autonomy which chooses the bad both is and is not valuable.

The whole spirit of Raz's work requires him to keep the claim that autonomy is valuable only in pursuit of the good. Therefore, he must either give up the claim that autonomy as such is valuable, or he must give up the claim that autonomy is autonomy even when it chooses the bad. Strange as it may sound, I think it is the last claim he should give up. He should admit that autonomy cannot choose the bad, or conversely, that one who chooses the bad is not (in that respect) autonomous. (Of course, admitting that autonomy is not autonomy when it chooses the bad also contracts the effective scope of the claim that autonomy as such is valuable. But the fundamental point is that autonomy cannot choose the bad.)

It would be much easier to see that autonomy cannot choose the bad (or, by the same argument, the worthless) if Raz repaired an omission in his list of the conditions of autonomy. He lists a number of capacities, both mental and physical, which are essential for autonomy. But I believe he never says specifically that one required capacity is the capacity to discern the genuinely valuable. (The closest he comes, perhaps, is an undeveloped reference to "one's emotional and imaginative make-up.") But surely this is a required capacity, if there are genuine values, and if autonomous choice is to be the kind of choice that shapes the moral life. I do not say one must discern the genuinely valuable unerringly. But one must be able to make judgments of this kind, and make them more often right than wrong. And if we once recognize that a

---

203. Id. at 381; cf. id. at 411-12 ("But while autonomy is consistent with the presence of bad options, they contribute nothing to its value.").
204. Id. at 417.
205. See, e.g., id. at 419.
206. Id. at 408.
capacity to discern the valuable is essential to autonomy, then it seems highly plausible that a failure correctly to discern, or a failure to be guided by one's discerning, is a failure of autonomy. In which case, autonomy is autonomy only when it pursues the good.\textsuperscript{207}

\textbf{Coda}

I have found a great deal in Raz to criticize. In case the reader has forgotten, let me reiterate my claim that Raz comes nearer the truth of political morality than anyone has for nearly a century. I would not trouble to disagree with Raz at such length, if I did not find much more to agree with.

\textsuperscript{207} One last, related, point on Raz's belief that we should not coerce people to avoid worthless or bad activities. Raz's version of the harm principle allows coercing people to prevent them from harming themselves. \textit{Id}. at 412-13, 419. He recognizes that one can harm a person, not merely by making him worse off than he is, but also by failing to make him as well off as he ought to be. \textit{Id}. at 415-16. (Nor can Raz limit this last claim by saying one harms a person only when one fails to make him as well off as he has a \textit{right} to be. This would commit Raz to rights to autonomy, since Raz thinks one can harm a person by failing to guarantee his autonomy. \textit{Id}. at 415-18.) Raz also recognizes that autonomy spent on worthless projects does not contribute to the value of one's life. Why then does not the person who spends himself on worthless projects harm himself by making his life less good than it ought to be? And why does this harm not justify coercion to avoid worthless projects? Surely it does, in principle. The upshot is that we see again how entirely the weight of Raz's argument against such coercion falls on the problematic claim that coercion is too clumsy a tool.
APPENDIX

IS THERE A THEOREM CALLING FOR EXCEPTIONLESS DEFERENCE TO AUTHORITY? (AND IF SO, IS THAT OBEDIENCE?)

The first question I pose—"Is there a theorem calling for exceptionless deference to authority?"—is raised by a tantalizing passage in which Raz argues that under certain conditions, the optimific decision rule for someone who can identify an authority over a certain class of cases is to follow that authority's directives in every case within the class. Raz states no theorem, nor does he carefully specify premises. But there is just enough of quasi-mathematical rigor in the argument to make one wonder whether it could be formalized. On the other hand, there is just enough of vagueness to make one wonder whether the argument is good at all.

Consider the following state of affairs. Jones has a stock portfolio. She must make repeated decisions about whether to sell blocks of shares. Past experience suggests that Jones makes the right decision in sixty percent of the cases she decides. Jones also has costless access to the opinions of a financial analyst who is able to identify the right decision in ninety percent of the cases. Clearly, Jones will do better in managing her portfolio if she always follows the analyst's advice than if she ignores the analyst completely and always relies on her own independent judgment. (I assume that the gains and losses from right and wrong decisions are the same in every case, or else that they are randomly distributed with respect to the rightness or wrongness of decisions by either Jones or the analyst.)

So, if Jones must either follow the analyst's advice all the time or follow her own independent judgment all the time, she is better off doing the former. But it occurs to Jones that maybe she can do better than either of these alternatives. For example, suppose that instead of following the analyst blindly, Jones always takes the case for doing what the analyst recommends to be fifty percent stronger than it appears to her in the exercise of her independent judgment (since the analyst's success rate is fifty percent higher than Jones's); but then, having given the reasons for the choice the analyst recommends this fifty percent extra weight, Jones still does her own balancing of the (partially re-evaluated) reasons favoring each choice. Obviously, there are a number of strategies like this which would allow Jones to take the analyst's advice into account without making it conclusive. Any such strategy will lead Jones to do
what the analyst recommends more often than she would if she ignored the analyst entirely. But it will leave some cases in which Jones still acts against the analyst's recommendation.

Raz wants to argue that in the situation as I have described it, any such strategy as we have imagined for Jones would be a mistake. The best course for Jones is to follow the analyst's advice in every case.²⁰⁸

Raz's argument goes as follows: If Jones uses any such strategy as we have imagined, she will end up doing what the analyst recommends in some number of cases. Specifically, she will do what he recommends in all those cases where her own uninfluenced judgment would have been the same as the analyst's, and also in those cases where her own uninfluenced judgment would have been different from the analyst's, but where the effect of feeding the analyst's view into her decision process is enough to tip the balance in favor of the course the analyst recommends. But in the remaining cases, where Jones's uninfluenced judgment would have been different from what the analyst recommends and where the effect of feeding in the analyst's view under her chosen strategy is not enough to tip the balance, Jones ends up acting against the analyst's advice. Raz asserts that: "If, as we are assuming, there is no other relevant information available,"²⁰⁹ then we can expect that in those cases where Jones does what the analyst recommends, her success rate will be the same as the analyst's, that is, ninety percent. But in those cases where she does not do what the analyst recommends, her success rate will be her success rate with her own independent judgment, that is to say, sixty percent. Obviously, Jones's overall success rate remains less than the analyst's success rate of ninety percent. Equally obviously, the only way for Jones to raise her success rate up to the analyst's is to follow the analyst's advice in every case. So the best course for Jones is to follow the analyst's advice in every case.

Now, this a superficially plausible argument. It seems that it must prove something. (And I trust the reader sees why it set me thinking about "theorems" on exceptionless deference.) On the other hand, the reader's antennae should have quivered when we came to the bit about there being no other relevant information available. Information is a notoriously slippery concept, and Raz says absolutely nothing to spell out what he means by his assumption that there is no other relevant information available. It might seem it is sufficiently obvious what that means, but I think it is not.

²⁰⁸. Id. at 67-69.
²⁰⁹. Id. at 68.
To bring out the first problem about Raz’s assumption, imagine that Jones, hearing Raz’s argument, responds as follows:

Your argument sounds good. I am not even certain what’s wrong with it. But I am going to go on using my chosen strategy all the same. Let me explain why. The analyst comes up with the right decision more often than I do in my uninfluenced judgment. But it does not follow that the analyst is right in every case in which I am right, and then some. There can perfectly well be some cases where my uninfluenced judgment is right and the analyst is wrong. Indeed, I think there probably are. I have not kept track of past experience closely enough to establish definitely what I am about to assert, but it seems to me a priori likely that my success rate (with my uninfluenced judgment) is a good deal higher in cases that seem to me very clear (that is, where the balance of reasons as I see them points very strongly in one direction) than in cases that seem to me close. I suspect that in cases which seem very clear to me, my success rate is even better than the ninety percent which is the analyst’s success rate over the entire relevant class of decisions. (Conceivably the analyst does even better than ninety percent in those cases which seem to him particularly clear, but I have no information about which those are.) What I am doing in effect, by the use of my strategy, is following the analyst’s advice except in those cases where his advice runs counter to my very strong judgments. This seems a reasonable course to me.

What does Raz say to this? First, can Raz say that Jones is definitely making a mistake, that her strategy is definitely inferior to the policy of following the analyst’s advice all the time? No, he cannot. Jones’s strategy might well be more effective than following the analyst’s advice all the time, and for just the reasons she offers. Unless Raz knows that Jones does not have an especially high success rate, using her own uninfluenced judgment, in those cases which seem to her very clear, he is not in a position to insist that Jones has elected an inferior procedure. (Actually, Raz recognizes this. When he is being careful, he does not claim that treating the analyst’s advice as preemptive must be the best policy. Rather, he says “we can expect” that it will be. I take this phrase out of the middle of Raz’s argument, not the conclusion, but it seems appropriate to regard it as qualifying the conclusion as well.)

If Raz cannot say Jones’s policy is definitely inferior in its results, can he retreat to the claim that Jones is nonetheless irrational when she follows it? Is it true that “we can expect” Jones to do better by following the analyst’s advice in all cases? Again, I think the answer is, No. Jones relies on an a priori assumption that she is especially likely to be right in cases where the correct decision seems to her very clear, but this seems a
perfectly reasonable assumption. I do not assert that in every kind of
case, we are more likely to be right when the case seems clear-cut than
when it seems close. One can imagine particular kinds of cases where
one is most likely to be wrong when the decision seems easiest. But I
suggest that on almost any model of decisionmaking which can account
for such success as people actually have in making decisions in their daily
lives, it will be true that in most kinds of cases, the likelihood of our
deciding rightly correlates positively with the degree to which the choice
seems clear. If that is so, then I do not see how Raz can say Jones is
being irrational when she prefers her chosen strategy to giving the ana-
lyst's advice preemptive force.\footnote{What precise strategy is most rational for Jones is an additional question, probably without
a definite answer. But that does not lead us back to full preemptive force.}

If Raz cannot say that Jones's policy is definitely inferior to the pol-
icy of following the analyst's advice all the time; and if he cannot even
say that in terms of the expected results of various policies, Jones is irra-
tional when she refuses to follow the analyst's advice all the time; what
can Raz say to rescue his argument? What he is compelled to say is that
if Jones is rational in relying on her strategy, then it is because she has
"other relevant information." The question now is, is that a plausible
ting to that, but it is worth noting.

To me, it seems implausible to say Jones is relying on "other infor-
man."

Jones is not relying on any proposition such as "The analyst is particu-
larly muddle-headed, and I am particularly astute, when the issue con-
cerns railroad stock," or "The analyst always does badly in March, when
he has just come back from vacation in the Bahamas and is out of
touch." (These are the sorts of propositions most immediately brought
to mind by Raz's reference to other information.) Nor does Jones rely on
claims that the authority can be seen to be clearly and incontrovertibly
wrong in certain particular cases.\footnote{Remember Raz's ambivalence about whether we should follow an authority when it is
clearly wrong. Whatever Raz thinks we should do in such cases, it is plain that the optimistic deci-
sion rule would require ignoring the authority in such cases, unless we expand the scope of the
discussion and bring in power-of-example effects and the like. So Raz's conclusion that the optimistic
rule is exceptionless deference cannot be right unless we exclude cases where the authority is clearly
wrong. So the knowledge that the authority is clearly wrong in some case must count as "other
information." I have no objection to that, but it is worth noting.}

Jones does not even rely on specific information about her own past performance in portfolio decisions. She
does not say, "I know from past experience that I do better than the
analyst in those cases which seem to me very one-sided." All Jones relies
on is a very general, but also thoroughly reasonable, assumption about
her decisionmaking in general. We can call that "other information" if we like, but it seems to be stretching a bit.

Suppose, however, that we decide to stretch the point and say Jones is relying on "other information." A new danger now appears, that Raz's conclusion will be degraded to a near-tautology. Even if we agree that Jones's strategy here depends for its justification on "other information," we still will not have given a general account of what is "other information" and what is not. We are edging in the direction of saying that "other information" is any information or any reasonable assumption that would justify Jones in using a decision rule other than "Follow the analyst in every case." But then the conclusion of Raz's argument becomes in effect "If Jones has no justification for not following the analyst in every case, then she ought to follow the analyst in every case." This is true, but not terribly interesting.212

Perhaps we can better deal with the questions about "other information" if we get a bit more formal. What the discussion to this point suggests is that if there is to be an interesting theorem requiring exceptionless deference, it will have to be something like the following:

If
1. Jones can identify some class of cases, \( C \), and an authority, \( A \), such that \( A \) does better on the whole at identifying what Jones should do in cases of the class \( C \) than Jones can do by her own independent judgment (reached without reference to the directives of \( A \)); and
2. Jones cannot identify by her own independent judgment any subclass, \( S \), of \( C \) in which she does better, by her own independent judgment, than \( A \) at identifying what she should do in cases of the subclass \( S \);

then Jones will achieve the best feasible results by following the directives of \( A \) in every case of the class \( C \).

(I trust it is clear that the second premise of this theorem is an attempt to capture the idea that Jones has no other relevant information. As we shall see, the attempt fails, in the sense that the theorem is not true. One way to describe the failure is this: It turns out that Jones may have no other information which is useful to her in the exercise of her

---

212. Note that we cannot at this point try to justify following the authority by reference to the costs of independent decisionmaking. The argument under discussion makes no reference to the costs of deciding, and to bring in such considerations would completely change the nature of the argument.
independent judgment, and yet have information which is useful in conjunc-
tion with the authority's directives and which suggests a course of 
action that is better than blindly following those directives.)

Notice that the example previously discussed, where Jones applies 
an "extra fifty percent weight" strategy, does not constitute a counterex-
ample to this theorem. If Jones in fact gets better results by her strategy 
than by following the analyst in every case, then she does not satisfy the 
second condition of the theorem. She is able to identify by her own 
independent judgment a subclass of cases where she does better than the 
authority, namely, those cases in which it seems to her (in her independ-
ent judgment) that the case in favor of the alternative she independently 
decides on is more than fifty percent stronger than the opposing case. So, 
we have dealt with the problem of Jones and the financial analyst. (Let 
me emphasize that we have dealt with the problem of Jones and the 
financial analyst only in the sense that we have got that problem outside 
the scope of our theorem. We have not in any way undermined the force 
of Jones's argument for the "extra fifty percent weighting" as an intui-
tively plausible argument for ignoring the advice of the analyst when the 
particular case seems to Jones, in the exercise of her own judgment, 
extremely one-sided.)

We have now stated a plausible sounding theorem, to which the 
story of Jones and the financial analyst is not a counterexample. Still, is 
the theorem true?

In the brief discussion that follows, I shall be quite conclusory. I 
shall make some assertions about when the theorem is true, and when it 
is not, without offering anything like rigorous proof. I have given some 
thought to this theorem, and I have constructed semi-formal arguments 
for all the claims I make, proving them to my own satisfaction. But 
actually setting out rigorous arguments would entail a complexity dispro-
portionate to the significance of the topic. In the course of just describ-
ing my conclusions, I shall say things which ought to get anyone else 
started who wants to think more about the theorem. And the most 
important point I shall eventually make, about why the theorem is not 
true in general, can be understood without any proof. That is to say, I 
shall make an observation which, without actually proving that there are 
counterexamples to the theorem, will make it sufficiently clear how the 
counterexamples, assuming they exist, arise.\textsuperscript{213}

\textsuperscript{213} The editors of the \textit{Southern California Law Review} take no responsibility for my unproved 
assertions. But I have consulted the Ghost of Fermat, who is willing to go on record that I have 
proofs for everything I say.
Returning to the theorem. Under certain conditions the theorem is true. First, the decisions Jones faces must be binary decisions—that is, there must be only two alternatives that Jones is choosing between. (Notice that this is a very substantial restriction.) Second, we must make assumptions that we might sum up by saying Jones is at least a "minimally competent decisionmaker." For example, we must assume that if Jones operates on her own independent judgment, she makes the right decision at least half the time (if she is faced with binary decisions). The intuitive justification for such an assumption is that if she makes the right decision less than half the time she would do better to randomize—and we assume in effect that if the best she can do is to randomize, she ought to be able to figure that out. The assumptions that characterize Jones as a minimally competent decisionmaker are not trivial; but without some such assumptions, I do not see how we could prove anything even remotely interesting. So, under certain fairly restrictive conditions, the theorem is true.

If we now abandon the restriction to binary choices—that is, if we allow Jones to confront choices where there are three or more options—the theorem is not true, even if Jones is a minimally competent decisionmaker. (The natural characterization of a minimally competent decisionmaker changes somewhat as we expand the kinds of choice situation that are possible, but we need not go into that.) Notice that if there are three or more options, we cannot force the problem back into a binary mode by saying the proper options to consider are simply "follow the authority" and "do not follow the authority." If there are three or more underlying options and the authority directs one, then there will be at least two ways of not following the authority, and the question whether it is better to follow or not to follow may have no determinate answer even in a single case.

Given that our theorem is not true in the general case, what is interesting is the reason it is not true. The reason, at bottom, is that the pronouncements of authority are themselves information. They are information which must become available to Jones before she can even consider following the authority. But once Jones has the information embodied in the authority's pronouncements, there may be better ways of using that information than just slavishly following the authority. For example, Jones might happen to know from past experience that, even though the authority does better than she does in general, still, when she and the authority disagree in some particular way, then her judgment is better than the authority's. What is crucial here is that the subclass of
cases in which Jones knows she does better than the authority may be a subclass which she cannot identify by her own independent judgment. She may know it, for example, only as the class of cases in which she chooses alternative number one and the authority directs her to alternative number three. (To say that Jones does better than the authority in this subclass is not at all to say that Jones does better whenever she chooses alternative number one. The authority may do much better than Jones when Jones chooses alternative number one and the authority chooses alternative number two.) So, there may be a subclass of cases where Jones does better than the authority, which she can identify only by comparing her own independent judgments with the judgments of the authority. If this is the situation, then Jones satisfies both conditions of our theorem, but it is not true that the best decision rule for her is to follow the authority in every case. In short, we have a counterexample to the theorem. (That this and other counterexamples I refer to really are logically possible is one of the things I have proved to myself by arguments more rigorous than I include here.)

Once we have the basic idea of this counterexample in hand, we can construct any number of variations. For example, Jones might know from experience that the authority does better than she does in general, but that whenever she (in her independent judgment) and the authority agree (which cases we assume to be rare), then they are both wrong, and what she should do is randomize over the other possibilities. (This could be the case even though Jones is a minimally competent decisionmaker, and even though we have assumed the authority does better than Jones in general.) In this situation, Jones can identify the relevant subclass (that is, the subclass where she ought to depart from the authority’s directive) only with the help of the authority’s directives (because the subclass in question is the subclass where she and the authority agree). And even in this subclass Jones does not do better than the authority in the exercise of her independent judgment (since both are always wrong). But still, there is a better decision rule for Jones to follow than exceptionless deference to the authority.

The point, as I have already observed, is that the judgments of the authority are information. They are information which Jones does not have (by definition) when she exercises her own independent judgment. But they are information she must acquire before Raz’s decision rule “follow the authority in every case” is available to her. And this information, which makes that decision rule available to Jones, also makes
available a number of other new decision rules, some of which may be better than exceptionless deference.\textsuperscript{214} It may seem that my counterexamples are bizarre. How likely is it, to take the more bizarre of the two examples I have sketched, that Jones and the authority should both be independently minimally competent, but that they should both be wrong whenever they agree? I concede at once that this is very unlikely. (Possible, I insist, but unlikely.) And perhaps all the possible counterexamples to the proposed theorem are almost equally unlikely.

Even so, we can learn from the bizarre counterexamples two things of importance. The less important of the things we learn is that the proposed theorem is false. That conclusion is not without intrinsic interest. And remember that even though each counterexample by itself may seem bizarre and unlikely, the counterexamples are a populous family. Having thought some about the theorem and the counterexamples, I am inclined to believe that there is no interesting theorem to be discovered here. I am inclined to believe that every true general claim along these lines is at best a thinly disguised variation on the near-tautology, “If Jones has no justification for not following the authority in every case, then she ought to follow the authority in every case.” But I could be wrong.

The much more important thing we can learn from the counterexamples has to do with the connection (or lack of necessary connection) between following authority and obeying it. The counterexamples compel us to notice that even after Jones has the information about the authority’s directives in hand, there may be better things to do with that information than just to follow the authority in every case. Once we understand that, we are led to understand a further point: \textit{Even if it turns out that the best thing Jones can do is to follow the authority in every case}, Jones is still just \textit{using} the authority’s directives in a particular way (as she might use them in a different way if the conditions were met of one of the bizarre counterexamples).

\textsuperscript{214} It may occur to the reader that what I have just said about the judgments of the authority as information is as true when Jones faces binary decisions as when she faces decisions with more than two options. So why is the theorem true when the decisions are all binary? It just happens that when the decisions are all binary, and when Jones is minimally competent (which also tells us something about the competence of the authority, since the authority does better than Jones across the board), then any decision rule which is better than “Follow the authority in every case” must be effectively available to Jones under some description even before she knows about the authority’s conclusions. In other words, in the binary case, if the conclusion of the theorem fails, so does the second premise. The binary nature of the decisions increases the strength of the combined assumptions that Jones is minimally competent and that the authority is better, and it also significantly limits the essential variety of possible rules.
Using the directives, even in a way which produces the same behavior as obedience, is not the same as obeying them. In other words, even if Jones decides the best course is to follow the authority's directives without exception, her relation to the authority is not one of obedience.\textsuperscript{215}

\textsuperscript{215} While this article was being edited, I happened upon a fascinating survey of research on the relative reliability of clinical (that is, particularized) and actuarial (that is, rule-guided) judgments in (mostly) medical and psychiatric contexts. R. Dawes, D. Faust, P. Meehl, \textit{Clinical Versus Actuarial Judgment}, 243 \textit{Science} 1668 (1989). (The issue of clinical versus actuarial judgment is not exactly the same as the issue of following one's own judgment or deferring to authority. But the parallel is close enough to make the study very interesting in the present context.) The study strongly suggests that most people, when offered information on judgments by some other source which they have reason to believe does better than their own independent judgment, even if they then invent some composite judgment process which combines the conclusions of their own judgment and that of the independent source, rely too much on their own judgment and too little on the other source, and end up doing worse than they would have if they had relied solely on the other source. In other words, they do worse than if they had practiced exceptionless deference. This, of course, does not undermine my claim that there is no theorem calling for exceptionless deference; and indeed the authors of the survey explicitly say that inadequate study has been done of the question whether an individual who was sufficiently conservative in her decisions to depart from the conclusions of the other source in favor of her own might not manage to improve on the other source. \textit{Id.} at 1671. But reading the survey reminds one powerfully of two truths we tend to forget: (1) human beings are not all that good as informal decisionmakers, and (2) we are strongly inclined to give ourselves credit for being a lot better than we are. Deference to wise authority has more to recommend it than we would like to admit.