



OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Just 'Cause You're Smarter than Me Doesn't Give You a Right to Tell Me What to Do:
Legitimate Authority and the Normal Justification Thesis

Author(s): Kenneth Einar Himma

Source: *Oxford Journal of Legal Studies*, Spring, 2007, Vol. 27, No. 1 (Spring, 2007), pp.
121-150

Published by: Oxford University Press

Stable URL: <https://www.jstor.org/stable/4494576>

REFERENCES

Linked references are available on JSTOR for this article:

[https://www.jstor.org/stable/4494576?seq=1&cid=pdf-
reference#references_tab_contents](https://www.jstor.org/stable/4494576?seq=1&cid=pdf-reference#references_tab_contents)

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<https://about.jstor.org/terms>



JSTOR

Oxford University Press is collaborating with JSTOR to digitize, preserve and extend access to
Oxford Journal of Legal Studies

Just 'Cause You're Smarter than Me Doesn't Give You a Right to Tell Me What to Do: Legitimate Authority and the Normal Justification Thesis

KENNETH EINAR HIMMA*

Abstract—Joseph Raz's famous theory of authority is grounded in three claims about the nature and justification of authority. According to the Preemption Thesis, authoritative directives purport to replace the subject's judgments about what she should do. According to the Dependence Thesis, authoritative directives should be based on reasons that actually apply to the subjects of the directive. According to the Normal Justification Thesis (NJT), authority is justified to the extent that subjects are more likely to comply with right reason by following the authority's directives than by following their own judgments about what right reason requires.¹

In this article, I consider a number of ways in which NJT might be construed as a justification for authority. First, I evaluate NJT construed as a principle that would provide a practical justification for an individual to accept or recognize a particular person or persons as a preemptive authority. Second, I evaluate NJT construed as a principle that describes the conditions under which a state or legal system is morally legitimate. I argue that NJT is true under none of these interpretations.²

1. *The Normal Justification Thesis*

The Razian analysis of authority begins with the idea that the conceptual function of authority is to 'mediat[e] between people and the right reasons that apply to them'.³ To discharge this function, a normative system, like a legal system, must

* Department of Philosophy, Seattle Pacific University. Email: bimma@spu.edu. I am indebted to an anonymous reviewer for a number of very helpful comments and suggestions.

¹ Joseph Raz, 'Authority, Law, and Morality' in Raz, *Ethics in the Public Domain* (Oxford: Clarendon Press, 1994) (hereafter referred to as *ALM*).

² I should point out that these interpretations of NJT need not be considered mutually exclusive. One might take the position that the proper interpretation of NJT varies according to context or that it functions in more than one way, perhaps, for example, as both a principle of practical rationality and a principle of moral legitimacy. I do not know whether these interpretations are correct, but it is not important for my purpose. If I am successful in showing that none of the specific interpretations succeed, then it follows that none of these interpretations of NJT provides the correct account of authority.

³ *ALM* 214.

satisfy two conditions. First, it must be the source of directives that reflect the balance of right reasons with respect to what subjects ought to do.⁴ Thus, according to the so-called Dependence Thesis, 'authoritative directives should be based, among other factors, on reasons which apply to the subjects of those directives and which bear on the circumstances covered by the directives'.⁵

Second, a normative system must, so to speak, stand between people and right reason in the following sense: the authority's decision must be able to replace (or preempt) the reasons that would otherwise be considered by the subject. As Raz puts it:

The [authority's] decision is ... a reason for action. They ought to do as he says because he says so ... [But] it is not just another reason to be added to the others, a reason to stand alongside the others when one reckons which way is better supported by reason ... The decision is also meant to replace the reasons on which it depends.⁶

According to the Preemption Thesis, 'the 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 what to do, but should replace some of them'.⁷

The 'service conception' of authority suggests a thesis regarding the justification of authority. According to the Normal Justification Thesis (NJT), authority is justified to the extent that the subject is more likely to do what right reason requires by following authoritative directives than by following her own judgment:

The normal and primary way to establish that a person should be acknowledged to have authority over another person involves showing that the alleged subject is likely better to comply with reasons which [objectively] 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, than if he tries to follow the reasons which apply to him directly.⁸

Though NJT is not implied by the service conception of authority, it coheres well with it: given the mediating function of authority, it is natural to suppose that authority is justified only insofar as it does a better job than its subjects of deciding what right reason requires.

To see the motivation for NJT, it is helpful to compare the justification for taking someone's advice. Consider a case in which one person A will be hurt if

⁴ Right reasons, for Raz, are those reasons that *objectively* apply to a person regardless of whether she is subjectively aware of them. Likewise, the balance of right reasons should also be construed as an objective notion: what the balance of right reasons requires does not depend on how the subject perceives that balance.

⁵ ALM 214.

⁶ ALM 212–13.

⁷ ALM 214. This distinguishes authority from other forms of normativity. Many normative propositions function as first-order reasons. For example, conditional normative statements, such as 'you should eat your broccoli to stay healthy', functions as a first-order reason that may compete with other first-order reasons, such as that expressed by the statement 'you should eat chocolate because it tastes good'.

⁸ ALM 214.

her friend B does not accept A's advice. The desire to spare A's feelings might, depending on the circumstances, be a reason for accepting A's advice; if the matter were sufficiently inconsequential and the advice was harmless, B might be justified in accepting A's advice to avoid hurting her feelings. But, as Raz points out, that is not the normal reason for accepting advice: 'The normal reason for accepting a piece of advice is that it is likely to be sound advice'.⁹ The normal justification for accepting advice, then, has to do with the quality of the advice.

The same is true, on Raz's view, for the justification of authority. There are many different ways a person might justify accepting authority. A person might, for example, accept the authority of a person as a way of defining her identity as a member of some group. Raz concedes that such identification can, depending on the circumstances, be 'a proper ground for accepting authority',¹⁰ but he argues that it is not the normal ground for doing so:

[T]he normal and primary way to establish that a person should be acknowledged to have authority 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.¹¹

Under NJT, then, authority is ultimately justified in terms of its comparative reliability in discerning what right reason requires: '[a]n authority is justified, according to the normal justification thesis, if it is more likely than its subjects to act correctly for the right reasons'.¹²

2. *Evaluating the Normal Justification Thesis*

A. *NJT as a Principle of Practical Rationality*

As it turns out, there are a number of ways to construe NJT. For example, Raz sometimes suggests that the relevant issue with respect to justifying authority has to do with the individual's acceptance of authority:

Identification is a common and often proper ground for *accepting authority*. It is therefore important to establish the reasons why it is no more than a secondary justification dependent on the availability, at least to a certain degree, of another justification. *Acceptance of an authority* can be an act of identification with a group because it can be naturally regarded as expressing trust in the person or institution in authority and a willingness to share the fortunes of the group which are to a large extent determined

⁹ Raz 'Authority and Justification', 14 *Philosophy and Public Affairs* (1985), 19. (hereafter referred to as *AJ*).

¹⁰ *AJ* 20.

¹¹ *AJ* 18–19.

¹² J. Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986), 61 (hereafter referred to as *MF*).

by the authority. But trust in an authority is trust that the authority is likely to discharge its duties properly. It therefore presupposes a principle which should govern its activities.¹³

This passage suggests that NJT should be construed as providing conditions that must be satisfied for an individual to be practically justified in *accepting* (or recognizing) a person or persons as a preemptive authority.

Construed as a principle of practical rationality, NJT purports to provide a reason that would justify a subject in accepting the preemptive authority of another person: acceptance of preemptive authority is justified if the authority's directives satisfy the conditions of NJT. Thus construed, a person is *practically justified* in recognizing or accepting preemptive authority if and only if she is more likely to comply with the demands of right reason by following the authority's directives than by following her own judgment about what right reason requires.

The task of evaluating this construction of NJT is complicated by the fact that there are at least two different classes of reasons that might be applicable in deciding what to do. To begin, practical rationality might require consideration of reasons—'prudential reasons'—that are concerned only with advancing or promoting the agent's own interests. For example, it might be rational to take a class because doing so will improve the agent's prospects on the job market. In this case, the applicable reason for taking the course is that the action conduces to what is the agent's self-interest.

More obviously, practical rationality might require consideration of moral reasons. For example, practical rationality might require a person who wants to break a promise to decide against doing so because it violates some morally protected expectation of the person to whom she made the promise. While prudential reasons might also dictate not breaking the promise (say, because it would result in retaliation or punishment), reasons having explicitly to do with conforming to moral rules are moral, rather than prudential, reasons.

On the construction of NJT being considered here, then, it provides either a prudential reason or a moral reason for accepting the authority of a person. In the following sections, I consider whether it is plausible to think that NJT functions as either a prudential or a moral reason for accepting authority.

(i) *NJT as prudential justification for accepting authority*

Construed as a *prudential* justification for accepting preemptive authority, NJT asserts the following: if A is more likely to achieve her self-interest by following the directives of B than by following her own judgment of what ought to be done, then A is prudentially justified in accepting B as a preemptive authority. For example, it is arguably in my self-interest to follow my doctor's advice to eat more fruit and vegetables because she is in a much better position than I am to

¹³ A7 20, emphasis added.

evaluate the balance of reasons as they pertain to my health. On this reading of NJT, then, I am prudentially justified in accepting my doctor's directive as a reason to eat more fruit and vegetables.

But notice that, construed as a prudential justification, NJT does not justify accepting my doctor's directive as a *preemptive* reason. There are two competing considerations here. The first is my prudential desire to give my doctor's directive due weight insofar as it increases the likelihood of attaining my self-interest. This consideration exerts practical force in the direction of giving the directive maximal consideration in deciding what to do. The second, however, pulls in the other direction. As Thomas May observes, acceptance of authority seems to preclude the exercise of autonomy:

Acting on what the authority judges ought to be done appears to circumvent one's own evaluational judgement, and thus autonomy. By circumventing the evaluational judgement of the subject it seems the subject is *prevented* from acting on her own determination of what ought to be done. The subject seems to be eliminated from the determination of her behavior.¹⁴

Acceptance of authority involves a commitment to comply with the authority's judgment regardless of whether it conforms to one's own assessment of the balance of reasons. And, thus, it requires giving up the prerogative, essential for the exercise of autonomy, to act on one's own assessment of the reasons.

In many instances, considerations of rational self-interest would seem to militate against surrendering the prerogative to act on one's own judgment; after all, no one has a greater interest in seeing me achieve my self-interest than I do. Accordingly, if the advantages of following my doctor's directives can be gained without giving up this prerogative, then considerations of rational self-interest favour doing so. Thus, insofar as I can secure the benefits of my doctor's directive by assigning it great weight, it seems to follow, as a prudential matter, that I should regard my doctor's opinion as having great weight but not as *preempting* my own judgment.

The notion of assigning great weight to a directive is, of course, metaphorical and hence somewhat vague. Indeed, it would be ideal to articulate precisely what is meant here since the notion of a weighing of reasons plays such an important role in many areas of jurisprudential theorizing.

However, this is not something that needs to be done here (and it is certainly not something that could be done here). In this connection, it is important to note that the idea of assigning great weight to some directive inherits much of its lack of clarity from the familiar idea—very common in the literature, including Raz's own work—of attempting to assess the 'balance of reasons' by weighing them. While the notion of 'great,' I suppose, adds some additional mystery to the notion, the real difficulty is in explaining in a theoretically rigorous way the

¹⁴ Thomas May, *Autonomy, Authority and Moral Responsibility* (Dordrecht: Kluwer Academic Publishers, 1998), 130. See also Heidi M. Hurd, 'Challenging Authority', 100 *Yale Law Journal* 1611 (1991).

weighing metaphor. In any event, I will simply assume that this can be done in a non-arbitrary way; if it cannot, then far more than my critique of Raz here is vitiated.

In any event, the claim that one can secure the benefits of following authority by assigning great weight to its directives suggests that NJT does not prudentially justify accepting preemptive authority. For the following seems to be a general principle of prudential justification:

Prudential Principle of Autonomy (PPA): If A can secure the advantages of following B's directives without loss of autonomy by assigning great weight to those directives, then A is prudentially justified in assigning them great weight—but not in regarding them as preempting A's judgment.

If PPA is a general principle of prudential rationality, then it operates to limit the scope of NJT in the following way. Construed in the light of PPA, NJT commonly functions, at most, as a reason to treat the directives of practical authority in the same way one treats the advice of a theoretical authority—namely, as having great weight but not necessarily preempting one's own judgment on the issue.

This, however, is not to deny that it can sometimes be rational for an agent to regard a directive as preemptive on purely prudential grounds. Whether or not this is the case depends on the distance between the ability of the authority in discerning what is in the best interests of the agent and the ability of the agent to do so. There are two limiting cases. First, if the agent knows that (1) an authority is perfectly infallible with respect to what is in the agent's best interests and (2) always sincerely directs what she takes to be in the agent's best interests, then the agent might be prudentially justified in accepting the authority's directives as preemptive—no matter how competent the agent is as long as she falls short of infallibility. Second, if the agent knows that she is perfectly incompetent with respect to evaluating her own self-interest (and hence with respect to evaluating the prudential merits of an authority's directives), then she might also be prudentially justified in accepting the authority's directives as preemptive—as long as the authority is more competent than the agent.

But this is not enough to rescue NJT as a principle of prudential justification. For the mere fact that recognizing preemptive authority may in special circumstances be justified on prudential grounds does not entail that recognizing preemptive authority can always, or even normally, be justified on prudential grounds if NJT is satisfied. To deny, for example, that the utilitarian principle functions as a first principle of morality does not deny that an act might sometimes be right because it promotes utility. Rather it is merely to claim that utility does not always, or even normally, provide a moral justification for behaviour.

What ultimately explains why recognizing preemptive authority might be prudentially justified where (1) the conditions of NJT are satisfied and (2) the authority is known to be better than the agent at discerning what is in the agent's self-interest is that the probable benefits associated with deference outweigh the probable costs associated with surrendering autonomy. But notice that (2) does

as much work as (1) in justifying the agent in recognizing the authority as a source of preemptive reasons. When (2) is satisfied, second-guessing the authority is very likely to lead to harm. If I know, for example, that my doctor is competent and that I am not competent to evaluate my doctor's advice, then it might be that I am not in a position reliably to secure the benefits of her opinion merely by giving it great weight. For the risk of allowing myself to second-guess her might dramatically increase the risk of making a potentially disastrous mistake. In such instances, it *might* be justifiable on prudential grounds to recognize my doctor as a preemptive authority. Since the antecedent of PPA is not satisfied, PPA does not pose a practical obstacle in such cases to accepting preemptive authority.

Now it might be that it is often the case that the antecedent of PPA is not satisfied for an agent because she is unable to secure the benefits of authority by assigning it only great weight. Perhaps situations commonly arise in which the epistemic distance between authority and subject is large enough that the subject cannot secure the benefits of following authority merely by giving its directives great weight. But the reverse also occurs often enough. Situations commonly arise in which the agent *can* secure the benefits of following authority without loss of autonomy by giving its directives great weight. While I am tempted to think that agents are not often in a position to competently second-guess authority, this is ultimately an empirical issue that cannot be settled here. This much, however, seems clear: if the situation in which an agent *can* competently second-guess authority is not the normal one, it is also not particularly unusual. For this reason, NJT does not appear to function as a prudential justification in normal circumstances.

In this connection, it is especially worth noting that, as an empirical matter, people do not behave the way one would expect if NJT functioned in normal circumstances as a prudential justification. Thanks to the widespread availability of medical information on the Internet, many people with serious illnesses take a proactive role in the management of their conditions. Patients with cancer, for example, often attempt to evaluate the advice they receive from their treating physicians against the most recent research. My suspicion is that this is most common among US patients with AIDS, who for certain demographic and social reasons have tended from the beginning to be more proactive, but it is a trend that is increasingly common with other diseases.

The prudential motivation here is straightforward. For many patients with a serious illness, the primary treating physician is a generalist who cannot keep up with research in one area. Given the severity of their conditions and the quality of available information, these patients should, as a prudential matter, evaluate their physicians' opinions. It may well be that in most cases of conflict the patient should accept her physician's recommendation. But this just requires that she assign great weight to the recommendation; it does not require her to regard it as preempting her own judgment.

Nor does Raz's view accord with how legal directives function in practical deliberations. When, as often occurs, a person jaywalks, exceeds the speed limit, litters, parks illegally, exaggerates deductions, etc., she cannot be treating the relevant legal directives as preemptive. Presumably what motivates noncompliance in these minor cases is the same thing that motivates noncompliance in serious cases: a judgment that it is in the agent's rational self-interest not to comply. And in many cases, it is clear that agents are prudentially correct in so behaving. A \$25 fine for illegal parking is a small price to pay, for example, to avoid losing one's job for being late.

Raz's theory, of course, is a conceptual theory and not intended to model the empirical behaviour of authority subjects. But that so many rational self-interested agents do not behave as one would expect if they were using NJT as a principle of prudential justification is a good reason for thinking NJT does not really function this way. For one would expect rational self-interested agents to have developed effective maximizing strategies. That NJT plays little role in the deliberations of agents who presumably have some sense for what is required by self-interest suggests that NJT is implausible as a principle of prudential justification.

(ii) *NJT as a moral justification for accepting authority*

As a moral justification for accepting or recognizing preemptive authority, NJT can be interpreted in two ways. First, NJT can be construed as describing the conditions under which accepting or recognizing preemptive authority is morally *obligatory*. Thus construed, NJT asserts the following: if 'the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) [by] accept[ing] the directives of alleged authority as authoritatively binding and tr[y]ing to follow them, rather than by trying to follow the reasons which apply to him', then the subject has a *moral obligation* to accept the authority as a source of preemptive reasons.

Second, NJT can be construed merely as describing the conditions under which accepting or recognizing preemptive authority is morally *permissible*. Thus construed, NJT asserts the following: if 'the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) [by] accept[ing] the directives of alleged authority as authoritatively binding and tr[y]ing to follow them, rather than by trying to follow the reasons which apply to him', then it is morally permissible, but neither good nor obligatory, for the subject to accept the authority as a source of preemptive reasons for action.

1. NJT as obligating principle

As a principle that defines a moral obligation, NJT is problematic because moral obligations seem to apply only to behaviour, and not to thoughts, feelings, or attitudes. A person can be morally obligated to provide emotional, intellectual, and financial support for her child—in other words, to *behave* as though she loves her child—but she cannot be morally obligated to *feel* love for her child.

For behaviour is subject to direct volitional control in a way that feelings are not: feelings just happen; they cannot be willed or willed away. And, as has been frequently observed, a person is not accountable for behavioural events not subject to direct volitional control. As the matter is sometimes put, 'ought implies can': the claim that *X* ought to do *a* implies that *X* can do *a*.¹⁵

It is, of course, reasonable to think that agents may have a moral obligation actively to cultivate certain sorts of feelings. A parent who lacks any real feeling for his child arguably has a duty to seek out any counselling or training that might help to promote a parental feeling in him. Though the agent lacks direct volitional control over whether he has the feeling, he exercises direct volitional control over whether he takes steps to seek out the appropriate help and thus can be said to exercise *indirect* volitional control over whether he has the feeling.

But the fact that feelings are often subject to indirect volitional control does not imply that a person can be morally accountable for her feelings. While a person can hence be held accountable for not taking appropriate steps to increase the likelihood of having a certain feeling, it is the failure to seek out help—and not the mental state itself—that is morally culpable. Strictly speaking, the failure to instantiate a feeling is not morally culpable because feelings are not subject to direct volitional control; and one can be morally obligated to do only what is within one's direct volitional control.

The same is true of beliefs. Whether an agent *A* believes a proposition *p* is not subject to direct volitional control. If the evidence that *p* is true strikes *A* as persuasive, she will believe that *p*. If the evidence that *p* is not true strikes *A* as persuasive, then she will believe that *not-p*. If the evidence strikes *A* as inconclusive and she has no psychological biases, she will likely be agnostic with respect to *p*. But how the evidence strikes *A* is not something over which she has direct volitional control. I could not, for example, choose to find inconclusive the evidence that the sun is shining today.

The reliability of an agent's beliefs is, of course, partly determined by her behaviour. If, on the one hand, the agent does little in the way of seeking out evidence, then she is more likely to accept a claim for psychological reasons unrelated to the likelihood that it is true. If, on the other, she actively seeks out evidence and argument, her beliefs are more likely to be reliable. And whether or not she actively seeks out such evidence is subject to direct volitional control—that is, if any behaviour is subject to direct volitional control. Accordingly, given that agents generally act on their beliefs, it seems reasonable to think that agents have a moral obligation actively to seek out the truth.

¹⁵ It is worth noting that Raz explicitly rejects this view: 'I am not a believer in "ought implies can"'. Joseph Raz, 'Facing Up', 62 *Southern California Law Review* (1989), 1153–1235, 1174. Raz is in a very small minority on this view; the principle that one cannot be obligated to do what one cannot do is at the very foundation of most normative and meta-ethical thinking. Given this, the burden of proof is on Raz to give a reason to reject this view.

But, as was true of feelings, while a person can thus be held accountable for having certain beliefs, it is not the mental state itself that is culpable. Rather it is the failure to do what is morally required to promote the relevant mental state—in this case, the failure to seek out the truth—that is morally culpable. Though an agent may have a moral duty to seek out evidence, it seems no more plausible to think an agent can be obligated to have certain beliefs than it is to think she can be obligated to have certain feelings. For if beliefs, like feelings, are not subject to direct volitional control, they are not subject to moral obligation.¹⁶

Of course, instantiation of an attitude accepting authority is neither a feeling nor a belief in any obvious sense. One can voluntarily resolve to treat someone's directives as preemptive reasons for action. Even so, such events are private not only in the sense that they are not publicly observable (i.e., they *cannot* be observed) but also in the sense that they seem beyond the reach of morality. While the arm of morality is long enough to reach public manifestations of private mental states, it does not seem to reach those mental states themselves.

It is, of course, true that we often make legitimate moral judgments about people based on their attitudes. I would certainly dislike and avoid someone who enjoys violent fantasies because I would judge her character as morally deficient. Nevertheless, my judgment is different from the judgment I would make if she harmed someone. I would regard her harming someone as blameworthy and as deserving of punishment, but not her enjoyment of the fantasy. My revulsion would be based on a judgment that someone who enjoys such fantasies is more likely to violate the moral obligation not to cause harm to others than someone who does not. While mental states are, of course, causally linked to behaviour, moral obligations apply only to behaviour. If this is correct, then NJT fails as an obligating principle because acceptance of authority, in the relevant sense, is a mental event that does not necessarily involve any public behaviour.

It turns out that Raz would agree that NJT does not imply a moral obligation to regard legitimate authoritative directives as preemptive. Raz, for example, rejects Hart's view that law provides *peremptory* reasons that 'preclude or cut off

¹⁶ One might think some beliefs are so objectionable it is simply wrong to have them, but it is hard to find a compelling example of such a belief. Consider the case of a racist who never acts on her discriminatory beliefs. Whether she violates any duties, I think, depends on the circumstances. Suppose, for example, that there is a possible world ω in which there are two continents, A and B, and two groups, reds and blues. And suppose the following propositions are true: (1) every red on A has greater intellectual ability and better moral character than every blue on A; (2) on A, reds and blues believe reds are 'superior'; (3) every blue on B has greater intellectual capacity and better moral character than every red on B; (4) on B, reds and blues believe blues are 'superior'; (5) on both continents, each person behaves exactly the way she would behave if she believed both groups are equal; (6) the average moral character and intellectual capacity of reds on A and B is the same as the average moral character and intellectual capacity of blues on A and B; and (7) no one knows there is another continent. My extremely reluctant intuition is that the benign racist in ω violates no moral duties. Of course, the actual world is very different from ω . A benign racist in this world violates a duty, but not simply by virtue of possessing racist views. We have a moral duty to seek out arguments and evidence that bear on our beliefs—particularly those that may express themselves in behaviour that affects other people. In this world, the benign racist must deliberately ignore the overwhelming evidence that her beliefs are false and hence fails to satisfy this moral duty. It is not so much the having of the belief that is culpable, but the failure actively to seek out truth—a duty that is owed not to herself, but to those people who can be affected by such beliefs. I am grateful to Steve Layman for suggesting this objection.

[in the law-subject] any independent deliberation or reasoning of his own'¹⁷ on the ground that the law does not even purport to obligate the subject to take a certain attitude towards valid legal norms—namely, as precluding deliberation. As Raz correctly points out, 'what counts, from the point of view of the person in authority, is not what the subject thinks but how he acts. I do all that the law requires of me if my actions comply with it'.¹⁸ Inasmuch as Raz believes that a subject satisfies all her obligations to the law by conforming her outward behaviour to it, he rejects the idea that NJT implies a moral obligation to regard authority as a source of preemptive reasons.

One might think that the notion of accepting authority, as Raz understands it, does not necessarily involve any sort of subjective internal state. On this construction of 'accepting authority', one accepts an authority if and only if one does what the authority says. Such a construction makes no reference to the internal subjective mental states of the agent. Since I will take myself to have shown that moral principles do not (at least not ordinarily) apply to these internal subjective states, I will assume this interpretation in the remainder of this section.

If this is what is meant by accepting authority, then the construction of NJT as an obligating principle is vulnerable to an objection that is similar to the one offered against the interpretation of NJT as a principle of prudential rationality. It seems clear, as a moral matter, that if one can secure all the morally-relevant benefits of following without giving up one's autonomy, then it is morally preferable to do so. And, again, it seems clear that one can achieve all this simply by assigning suitably great weight to the authority's directives.

2. NJT as permissive principle

One might construe NJT as a permissive principle of moral justification. Consider, for example, what Raz has to say about 'respect for law', an attitude that might be partly constituted by accepting law as preemptive authority:

[T]here is an attitude to law, generally known as respect for it, such that those who have it have a general reason to obey the law, that their reason is their attitude, the fact that they so respect the law, and that it is morally permissible to respect the law in this way (unless it is a generally wicked legal system).¹⁹

If respect for law is constituted by acceptance/recognition of law as a source of preemptive reasons, then NJT functions as a permissive standard for accepting/recognizing authority in the following way: it is *morally permissible* to accept/recognize authority as providing preemptive reasons for action whenever the conditions of NJT are satisfied.

¹⁷ H.L.A. Hart, *Essays on Bentham* (Oxford: Clarendon Press, 1982), 253.

¹⁸ *MF* 39.

¹⁹ *AL* 250.

There are a couple of problems with this construction of NJT. First, the claim that an act *A* is morally permissible does not function in practical deliberation as a reason for doing *A*. Insofar as the permissibility of *A* presupposes the permissibility of not-*A*, the permissibility of *A* indicates the absence of a moral reason for not doing *A*; otherwise put, the permissibility of *A* implies that morality is agnostic with respect to whether *A* should be done. It does not provide a reason that pulls in one direction or the other with respect to the desirability of doing *A*.

Second, NJT assumes that it is presumptively impermissible for a person to accept an authority as a source of preemptive reasons; otherwise, there is no need for a justification standard. But insofar as an act is presumptively impermissible, it is because it violates a presumptive moral obligation. Accordingly, construing NJT as a permissive principle presupposes that a person has a presumptive moral obligation not to accept preemptive authority.

To whom would such an obligation be owed? It seems implausible to think that one person owes a presumptive moral obligation to *others* not to accept authority. There are, of course, special circumstances that give rise to such obligations; a father, for example, owes a moral obligation to his child not to accept just anyone's authority. Nevertheless, such circumstances are the exception and not the rule; in the absence of a special relationship between A and B, there is little reason to think A is presumptively obligated to B not to accept authority. If a person has such an obligation, it is one she owes to herself.

The idea that we owe moral obligations to ourselves is an unnatural one. Though we often say things like 'John owes it to himself to get an education', we are more likely to intend them as prudentially normative than as morally normative. Thus, in most cases, they are more plausibly interpreted as statements of the form 'John is obliged to do X for the sake of self-interest' than as statements of the form 'John is obligated to do X for the sake of morality'.

This, however, is not to deny that we can make sense of such statements as expressing morally normative claims—or that people sometimes intend them to express such claims. There is a sense in which one can reasonably claim that, other things being equal, not getting an education or not taking more time with a careful decision has negative moral value. On this usage, one can thus reasonably claim that John morally should get an education or that John morally should take more time in deciding.

But the operative 'ought' is weaker than the 'ought' of obligation. 'Ought' can be used to recommend acts that go beyond the call of duty; thus, for example, one might remark 'John ought to give more to charity this year' without saying something that implies that he has not satisfied what obligations he has to give to charity. Similarly, 'ought' can be used to recommend the development of virtuous character traits; one might state, for example, 'John ought to be more sensitive to the effects of poverty'. Insofar, then, as we interpret 'John owes it to himself to get an education' as morally normative, the operative 'ought' falls short

of expressing an obligation and is not the sort of thing that is *owed* to someone who thereby has some sort of claim for satisfaction.

If ordinary usage suggests people do not often morally wrong themselves, it remains nonetheless reasonable to think that it is *possible* to do so. Committing suicide or inflicting grievous bodily harm on oneself might be examples of such moral wrongs. But even here the idea that a person owes herself a moral obligation not to kill herself seems problematic when contrasted with the more natural idea that she owes others a moral obligation not to kill them. It seems paradoxical, if not incoherent, to think that a person who harms herself in violation of a moral obligation owed to herself should be blamed or punished.

In any event, this much seems clear: assuming we owe ourselves moral obligations, they involve only activities that would create a risk of great harm to ourselves. Such obligations, if any, likely go no further than the obligations we owe to other people. Just as we are obligated to refrain from killing or causing grievous bodily harm to other people, so too, on this view, would we be obligated to refrain from killing or causing grievous bodily harm to ourselves. While it might be the case that we ‘ought’ to better ourselves in various ways, the ‘ought’ simply does not rise to the level of an obligation because the failure to do so does not involve a great risk of harm.

If this is correct, then the idea that accepting authority is presumptively wrong, which is entailed by NJT construed as a permissive principle, presupposes that accepting authority creates a risk of great harm to the subject. While accepting a person’s directives as binding can have disastrous consequences in some contexts (e.g., medical contexts), such contexts are comparatively unusual. There are many contexts in which I accept authority for no better reason than convenience precisely because there is no real risk of injury to my interests.

One might nonetheless think that acceptance of authority *necessarily* involves a great harm insofar as it results in surrendering autonomy, and that is why a *moral* justification is needed. But even if this is correct, NJT cannot morally justify accepting preemptive authority for the same reason it cannot prudentially justify doing so. As long as the benefits of accepting authority can be achieved without giving up the prerogative to act on one’s own assessment of the reasons, moral considerations seem to militate against accepting authority as a source of preemptive reasons.

This construction of NJT, then, fails as a permissive principle of moral justification—at least insofar as Raz intends NJT to justify ordinarily competent persons in accepting preemptive authority. While there may be instances where the distance between the agent’s capacity to discern the requirements of right reason and the authority’s capacity is large enough morally to justify the agent in accepting the authority as a source of preemptive reasons, this will be the exception and not the rule. Such cases justify acceptance of preemptive authority because the harm in surrendering autonomy will be outweighed by the harm in

not doing so. But in the vast majority of cases in which reasonably competent agents are involved, NJT will fail as a permissive principle.

B. *NJT as a Principle of Moral Legitimacy*

(i) *Construing NJT as a moral principle: a second source of ambiguity*

NJT need not be construed as a moral principle that justifies subject acceptance of preemptive authority. Instead, it can be construed as an account of moral legitimacy. On this construction, NJT distinguishes authorities that are morally legitimate from merely *de facto* authorities that are not morally legitimate. Given that the legitimacy of an authority is usually associated with its having a 'right to rule,' this construction of NJT would distinguish authorities that have a right to rule from merely *de facto* authorities that do not have a right to rule.

These two different kinds of account are not logically equivalent. Morally justified acceptance or recognition of authority by an individual does not entail that the authority has a 'right to rule' that individual. The idea that it is rational for an individual to accept preemptive authority to the extent she is more likely to comply with right reason by following the authority's directives than by following her own judgment has application to other forms of authority than just state authority. Thus construed, NJT would, if sound, justify accepting a doctor's authority as a source of preemptive reasons. But *qua* legitimacy principle, NJT would not imply the existence of anything resembling a 'right to rule,' even a restricted one, on the part of a doctor; nor would it give rise to a moral duty on the part of the patient to comply with the doctor's directive. Morally justified acceptance of an authority, by itself, does not entail a right to rule on the part of that authority.

Conversely, it might be that the conditions for legitimacy do not require practically justified acceptance of authority as a source of preemptive reasons for action. Under such circumstances, a legitimate authority would have a right to rule subjects even if it is not rational to regard those reasons as providing preemptive reasons for action. Heidi M. Hurd, for example, argues it can never be practically rational to accept preemptive authority because it violates the principle that an agent should always act on the balance of reasons available to her.²⁰ Insofar as the Razian account of authority requires an agent to comply with an authoritative directive regardless of whether the directive conforms to the balance of reasons, it violates this central principle of rationality by requiring the agent to ignore reasons that would otherwise apply to her. But assuming that Hurd is correct in thinking that acceptance of preemptive authority is never practically rational, it does not obviously follow that all ostensibly preemptive authority is *morally* illegitimate.

While the idea that authority could be morally legitimate without being rational may feel somewhat paradoxical, it is not obviously problematic. The

²⁰ Heidi M. Hurd, 'Challenging Authority', 100 *Yale Law Journal* 1611 (1991).

plausibility of such a view will depend on what exactly legitimacy is supposed to buy in the way of a 'right to rule'. If what is meant here is just that it is morally permissible for the authority coercively to enforce its directives, this might imply that it will be rational, at least in most cases, *to comply* with the directives of morally legitimate authority. But there is no obvious reason to think that it must be rational to accept the authority's directives *as preemptive*. As unnatural as the notion of such a state of affairs may seem, there is nothing in the notion of moral legitimacy that obviously precludes it.

Merely showing that NJT fails as a justification for recognizing authority, then, does not suffice to show that NJT fails as a moral justification of authority; for it might be that NJT succeeds as an account of moral legitimacy despite failing as a principle of practical justification. For this reason, a full evaluation of the Razian conception of authority requires addressing the question of whether NJT succeeds as an account of moral legitimacy.

(ii) *NJT and the problem of legitimate state authority*

Though, as we have seen, Raz commonly writes in terms of individual recognition of authority, he seems more often to intend NJT as an account of legitimate authority:

[E]xercise of authority involves a claim that those subject to it have a duty to obey it. This raises the special problem of the legitimacy of authority ... What can justify holding some people to be duty-bound to obey others? Under what conditions can some have a right to rule others?²¹

Later, Raz offers NJT as an answer to these questions: 'authority is justified only if its directives make it more likely that those subject to them will, if they acknowledge the legitimacy of the authority, better conform to reasons which apply to them'.²²

This sort of remark is common in Raz. For example, Raz describes NJT as 'a moral thesis about the type of argument which could be used to establish the legitimacy of authority'²³ and locates NJT among other theories of moral legitimacy:

As John Bell pointed out to me, if there is a common theme to liberal political theorizing on authority it is that the legitimacy of authority rests on the duty to support and uphold just institutions, as, following Rawls, the duty is now called. But that duty is of course dependent on a prior understanding of which institutions are just. The account here offered is meant as a beginning of an answer to that question.²⁴

²¹ Joseph Raz, 'Government by Consent' in John W. Chapman and J. Roland Pennock (eds), *Authority Revisited* (New York: New York University Press, 1987), 76 at 77–8 (hereafter referred to as *GC*).

²² *GC* 79.

²³ *AJ* 18.

²⁴ *AJ* 28.

And, on Raz's view, the legitimacy of authority turns on whether it has certain moral properties: 'If [a legal system] lacks the *moral* attributes required to endow it with legitimate authority then it has none ... To claim authority it must be ... a system of a kind which is capable in principle of possessing the requisite *moral* properties of authority'.²⁵ The frequency of such passages suggest that Raz's principal concern, as far as justifying authority is concerned, is with the moral legitimacy of state authority.

The problem of state legitimacy arises, on Raz's view, because authoritative directives purport to preempt the subject's own judgment and are generally backed with the police power of the state. As Raz aptly puts the matter:

A government does not merely say to its subjects: 'Here are our laws. Give them some weight in your considerations. But of course you may well be justified in deciding that on balance they should be disobeyed.' It says: 'We are better able to decide how you should act. Our decision is in these laws. You are bound by them and should follow them whether or not you agree with them'.²⁶

The need to justify state authority arises, then, because the state enforces its directives with the kind of coercive force that is clearly impermissible in nearly every other context.²⁷ While state authority is one instance of preemptive authority, the centrality of coercive enforcement mechanisms to legal practice makes the problem of justifying state authority especially acute.

It is difficult to see how it could be morally permissible for the state coercively to enforce a law (by means of incarceration—or worse) in the absence of some sort of general moral obligation to obey it. It seems to be a straightforward ethical truth that it is *morally* permissible coercively to enforce a rule against an agent only to the extent that the agent is *morally* culpable for violating the rule. Further, it seems to be a conceptual meta-ethical truth that an agent *A* is morally culpable in doing ϕ if and only if the performance of ϕ violates some sort of obligation owed by the agent not to do ϕ . If these two claims are correct, then it is a necessary condition for the permissible enforcement of a rule against an agent that the rule give rise to a moral obligation on the part of the agent to do what it requires. Any adequate account of legitimacy, at the very least, would hence

²⁵ *ALM* 215, emphasis added.

²⁶ *GC* 81.

²⁷ Strictly speaking, this passage does not say anything about the coercive enforcement of the law by authorities; it explicitly mentions only the claims of a government. There is little reason, however, to interpret this passage as being concerned only with the claims of government. What a government *claims* is a comparatively minor problem from the standpoint of political morality compared to what a government *does* by way of coercively enforcing laws against citizens. I might, I suppose, worry a little bit about whether these claims are correct and whether the government is thus morally justified in making them. But I am far more concerned about whether a government is justified in subjecting persons who do not comply with the law to incarceration (or worse)—and here it should be remembered that every court order, including those made in the context of civil disputes, is backed with the contempt sanction, which allows incarceration. The problem of justifying the claims of government (as opposed to these practices), as far as I can tell, has attracted little, if any, interest among political and legal philosophers. For this reason, I think Raz is best construed here as being concerned with the enforcement of law. In any event, this is how I will interpret him in what follows.

have to imply that there is some kind of moral obligation to obey the directives of a legitimate authority—at least those that are permissibly enforced.²⁸

Surprisingly, more appears needed to ground the permissibility of punishment than just a moral obligation to obey because the existence of such an obligation does not, by itself, imply the permissibility of using the state's police power to enforce it. The mere fact that I am morally obligated to do *X* implies that I am morally culpable for not doing *X*, but it does not imply that it is permissible for any particular person or persons to enforce that obligation against me.

Uncontroversial examples of legitimate coercive authority are very difficult to come by, but parental authority is a familiar example that seems to provide some light on the issue. Although a 16-year-old child has many *moral* obligations, not everyone can permissibly enforce those obligations against that child. In many instances, it would be wrong for anyone but the child's parents to attempt to punish the child. Thus, the claim that a person has a moral obligation to do *X* does not, by itself, logically imply that it is permissible for any particular party, including the state, to enforce that obligation.

What ultimately explains the coercive authority of the parent is that the child owes an obligation to the parent. A child's parents may use coercive means to teach proper behaviour to a child who is too young to be morally accountable for her behaviour, but this, strictly speaking, does not constitute punishment. When in the course of a child's development it becomes appropriate to *punish* her for disobeying her parents, it is partly because the child has, at least in legitimate cases, a moral obligation to obey her parents that is *owed to* the parents.

But notice that being owed a moral obligation, by itself, is not sufficient to explain the permissibility of punishment. If someone else's child steals from me, the child has breached an obligation that is owed *to me*; but it would, in ordinary circumstances, be wrong for *me* unilaterally to punish the child. What makes it permissible, on traditional views, for the child's parents to punish (as opposed to teach or correct) in this situation is (1) that the child violated a *content-independent* obligation to do what her parents have told her to do; and (2) that this *content-independent* obligation is owed to the parents. In ordinary cases (i.e., those where the parents are meeting their moral obligations as parents), then, the child's failure to obey, in and of itself, gives rise to a wrong against the parents that can be legitimately punished by the parents. What explains the permissibility of parental punishment, then, seems to be that the parents are owed a content-independent obligation of obedience by the child.

Exactly the same relational property seems to be necessary to explain the permissibility of state coercion. Insofar as the legitimacy of a state entails that it is *generally* permissible for the state to punish subjects for violating its directives, it is commonly thought that its legitimacy must imply that the failure to obey those

²⁸ This does not imply the existence of a general obligation to obey the law in legitimate systems—only an obligation to obey laws that are permissibly enforced.

directives, regardless of content, is morally culpable. There are, for example, many different ways a state might regulate, say, the flow of traffic, but it is permissible for a legitimate state coercively to enforce its directives with respect to the flow of traffic no matter which one it chooses. This, of course, is not to deny that there may be content-based constraints on legitimacy; a state that, say, requires discrimination against a particular ethnic group in employment and housing cannot be legitimate. But assuming there are content-based constraints on legitimacy, once those constraints are satisfied, state directives give rise to obligations that are independent of the content of those directives (as long as they fall within the scope of the authority's legitimacy). Accordingly, the moral obligation to obey the directives of legitimate authority is generally thought to be *content-independent*—though a directive's content may also give rise to an obligation, as when a legitimate authority prohibits theft.

The analogy between parental and state authority will only go so far; however, it is important to note that there are few philosophical resources for addressing the problem posed by state coercion. Apart from the existence of a perfect God (if such there be), parental authority is the only example to my knowledge of a coercive authority whose legitimacy is less problematic than that of state authority. For this reason, it is one of the few heuristic resources that could help us understand the conditions under which the directives of a putative authority may be coercively enforced. While legitimate state authority obviously differs in many respects from legitimate parental authority, it seems reasonable to think that what explains the permissibility of administering punishment in the latter case might also explain it in the former. If this is correct, then the parental example shows that an authority is justified in coercively enforcing her directives only if (1) the subjects have a content-independent moral obligation to obey the directives of the authority (2) that is owed *to* the authority.

The idea that the source of a directive could, by itself, give rise to an obligation to obey has not been universally accepted,²⁹ but this is the way that Raz thinks of legitimate authority. He points out, for example, 'No system is a system of law unless it includes a claim of legitimacy, of *moral* authority. That means that it claims that legal requirements are *morally* binding, that is that legal obligations are real (*moral*) obligations arising out of the law'.³⁰ To *claim*, then, that a state is legitimate is tantamount to *claiming* that there is a general content-independent moral obligation to obey law because, on Raz's view, there is a general content-independent moral obligation to obey the directives of a legitimate state.

²⁹ See, e.g. William A. Edmundson 'Legitimate Authority without Political Obligation', 17 *Law and Philosophy* (1998), 43. On Edmundson's view, to say that an authority A is legitimate over a person P is to say that P has a moral obligation to refrain from interfering with A's efforts to 'administer' her directives. See also David Copp 'The Idea of a Legitimate State', 29 *Philosophy & Public Affairs* (1999), 3–45.

³⁰ Joseph Raz, 'Hart on Moral Rights and Legal Duties' (1984) 4 *OJLS* 123–131 (emphasis added).

Indeed, in classically Hohfeldian fashion, Raz conceptualizes the notion of a right to rule as a logical correlative of an obligation to obey that is owed to the right-holder:

[Authorities] claim ... a right [to rule], i.e., they are *de facto* authorities because they claim a right to rule as well as because they succeed in establishing and maintaining their rule. They have legitimate authority only if and to the extent that their claim is justified and *they are owed a duty of obedience*.³¹

Moreover, on Raz's view, this moral obligation is content-independent: it is 'the fact that a [legitimate] authority requires performance of an action'³² that makes that performance morally obligatory; thus, he concludes that 'authoritative reasons [are] content independent'.³³

It is true, of course, that Raz is sceptical that there are any societies in which there is a 'general obligation' to obey law:

[T]he authority which states and governments claim cannot be based on the main argument for the justification of authority, i.e., that described by the normal justification thesis. Even reasonably just states claim more extensive authority than they are entitled to by that criterion ... [This argument, then,] denies the existence of a general obligation to obey the law even in a reasonably just society.³⁴

The problem here is that even reasonably just societies routinely exceed the scope of their legitimate authority and hence the scope of their ability to create content-independent moral obligations.

But, as the passages above make clear, Raz believes that directives that fall within the scope of a state's legitimate authority give rise to content-independent moral obligations to obey. Thus, as Raz puts the point, 'what is validly required by legitimate authority is one's duty, even where previously it was merely something one had sufficient reason to do'.³⁵ As an interpretive matter, then, it is clear that Raz believes that satisfaction of the conditions for legitimacy by a state entails that its directives give rise to a content-independent moral obligation to obey – at least within the scope of the state's justified or legitimate authority.

(iii) *NJT as sufficient condition for moral legitimacy*

Construed as an account of legitimate authority (as Raz understands that notion), NJT distinguishes legitimate authorities whose directives give rise to content-independent moral obligations from *de facto* authorities whose directives do not. On this construction, A's directives give rise to content-independent moral obligations to obey if and only if the putative subject is more likely to comply with the demands of right reason by following the directives of A than

³¹ MF 26, emphasis added.

³² MF 46.

³³ MF 35.

³⁴ MF 70.

³⁵ MF 60.

by following her own judgment about what right reason requires. Thus construed, NJT purports to provide sufficient and necessary conditions for moral legitimacy.

In what follows, I argue that NJT fails to state a sufficient condition for moral legitimacy. First, satisfaction of NJT does not entail that the directives of legitimate authority give rise to content-independent moral obligations to obey and hence does not guarantee the permissibility of coercive enforcement, which is the focal point of the problem of legitimacy. Second, construed as a sufficient condition for legitimacy, NJT is inconsistent with there being any necessary procedural constraints or substantive constraints on the scope of legitimate authority.

1. NJT, the obligation to obey, and coercive enforcement

NJT is problematic as an account of moral legitimacy because satisfaction of NJT does not suffice to give rise to content-independent moral obligations to comply with the directives of an authority. Suppose that right reason demands always that we comply with moral standards. Suppose further that A is infallible in determining what is required by moral standards and is morally impeccable in the sense that A always directs what is required by those standards. And suppose that there is no other fact that might legitimate A's authority apart from A's impeccability. While I am morally obligated to comply with the correct moral judgments that are expressed in A's directives, it is false that I am morally obligated *by* A's directives. If A's directives always demand what is required by morality, I am morally obligated to do as A says—but not for the reason that A says it.

The reason I am morally obligated to conform to A's directives is *content-dependent*. By hypothesis, the content of A's directives reflect the content of moral principles that already bind me; it is the content of *those* principles that give rise to pre-existing moral obligations that are, so to speak, simply repeated in A's directives. The fact that the source of A's directives is A adds nothing to that obligation to obey. Thus, whereas I am morally obligated to conform my behaviour to what A directs because A's directives demand what I am already morally obligated to do, A's impeccability alone cannot give rise to a *content-independent* moral obligation to *obey* A.

This is not, however, to deny that an authority can change a person's reasons for action. As Raz observes, the claim that authorities can make no difference with respect to the reasons people have (the 'no difference thesis') is problematic because '[i]t fails ... to explain the role of authority in the solution of coordination problems'.³⁶ Authorities can change the reasons people have for action by issuing a directive that solves a coordination problem—like one that requires people to drive on the right side of the road, rather than the left.

³⁶ MF, 30.

It seems clear, of course, that one has a moral obligation to obey such directives and that the fact that an authority has issued it changes the reasons people have with respect to which side of the road they should drive on. If this amounts to a situation in which one has a moral obligation to obey this directive because of its source, it should be clear that this is exceptional.

But I do not think that this is such a situation. As far as my intuitions are concerned, someone who drives down the wrong side of the street has committed a serious moral breach, but not because she has disobeyed the authority that issued the directive. The reason she has done wrong is that her act endangers lives because people will generally comply with the directive—something that is presumably known (or should be known) by the driver who has violated this directive. As far as I can tell, the fact that the directive has a particular source is morally irrelevant except insofar as other people will conform their behaviour to directives that come from this source—which is the real issue.

It is true, of course, that the fact that the authority has issued this directive changes people's reasons for action, but not by virtue of making disobedience itself a wrong-making property of an act. Rather, the authority has changed people's reasons for action because others will rely on the authority to settle coordination problems, where the key concern for everyone is to ensure that everyone's behaviour agrees. In any event, it should be clear that there is no content-independent moral obligation to obey the directives of a legitimate authority.³⁷

This result gives rise to two related problems with respect to A's authority. First, while it might be permissible for someone (perhaps God) to use coercive means to ensure that I do what morality requires, it is not morally permissible for any person to use coercive means against me to ensure that I *obey A*. Again, since A is morally impeccable, compliance with morality's demands entails compliance with A's directives. But if I fail to do what A directs, I am culpable because I failed to do what morality requires and not because I failed to do what A directs. If moral culpability is a necessary condition for justified use of coercion, then *no one* may use coercive means against me to ensure my *obedience to A* because *disobedience of A* is not in itself culpable.

Second, there is nothing that would ground any legitimate claim on the part of A coercively to enforce any moral obligations I might have. As we have seen, the claim that disobedience of a standard is morally culpable does not, by itself, entail that it is permissible for any particular person A to use coercive means to enforce compliance with that standard; it seems to be a necessary condition for A

³⁷ There is another way to see this point. Philosophers (including theists) have almost universally rejected the divine command theory of morality on the ground that it falsely presupposes that God can create content-independent moral obligations to act. The now-standard objection is that if God manufactures morality, then it follows that God can bring about a world in which people are morally obligated to torture infants for fun, something that is necessarily impermissible. But if it is implausible to think that an all-perfect God lacks a general capacity for creating content-independent moral obligations to obey, it is all the more implausible to think even the best finite authority figures can do so.

legitimately to punish noncompliance that it breaches an obligation owed to A. But even if it is false that A may justifiably punish B only for breaches of obligations owed to A, something more than mere impeccability is needed to justify A's punishing B.

Accordingly, the comparative reliability of the state in issuing directives that reflect the balance of reasons is insufficient to justify *the state* in imposing punitive sanctions for disobedience; since even impeccability with the requirements of right reason by a person does not, by itself, entail that it is permissible for that person coercively to enforce her directives, it follows that satisfaction of NJT by the state does not, by itself, entail that it is permissible for the state to enforce its directives.

2. NJT is inconsistent with procedural constraints on legitimacy

There are other problems with this construction of NJT. Construed as a sufficient condition for legitimacy, NJT implies that there are no other limits on legitimate authority. On many traditional accounts of legitimacy, this is problematic because there are necessary procedural constraints to moral legitimacy. According to these accounts, respect for subject autonomy requires that subjects be allowed to participate in a meaningful way in making decisions that vitally impact their well-being and hence requires decision-making procedures that are at least minimally responsive to citizen input. This, of course, does not imply that such procedures must be democratic in character. But what it does seem to imply is this: no matter how reliable an authority's judgment might be, a regime that fails to provide any procedural mechanism to address citizen grievances seems to fall well short of moral legitimacy.

To see the problem here, it will be helpful to consider two imaginary legal systems I will call Coercia and Consentia. Suppose that the citizens of Coercia and Consentia often make mistakes about the requirements of right reason; assume that they are comparable in this regard to the citizens of any existing country. Suppose, further, that each citizen of Coercia is neither more nor less likely than any citizen of Consentia reliably to discern the requirements of right reason. The point, then, of these two assumptions is to ensure that the ability of any given subject in one of the two regimes to discern the requirements of right reason is both ordinary and roughly equal to that of any other citizen in both regimes.

The next set of suppositions defines the relevant properties of the officials. Suppose that Coercian officials are impeccable in the sense that the laws they enact make obligatory all and only behaviours required by right reason. The impeccability of Coercian officials, then, entails, for all acts *a*, there exists a valid norm *L*, such that *a* is required (or prohibited) by right reason if and only if there is a legal obligation under *L* to do (or refrain from doing) *a*. Impeccability, then, implies infallibility about the requirements of right reason. In contrast, suppose the ability of Consentian officials to discern the requirements of right

reason is quite ordinary; indeed, they are no more reliable in this regard than Consentian citizens. The ability of any given Consentian official is thus equal to the ability of any subject in either regime.

A few assumptions about the way in which the two regimes rule will complete the picture. Suppose that officials in Consentia are democratically elected and hence enjoy the support of the subjects, who retain the power to remove them from office. Consentian officials provide a wide variety of mechanisms for citizens to address the state and they always actively solicit—and genuinely consider—the views of citizens before making decisions on matters vital to their interests. In contrast, the Coercian officials wrongly believe that the subjects belong to an ‘inferior race’ and cannot adequately govern themselves. Thus, while Coercian officials are impeccable, they have nothing but contempt for the Coercian citizens and never even pretend to pay attention to their views. In consequence, the subjects in Coercia despise the officials and long for, at the very least, a system minimally responsive to what they say. So consuming is this hatred that they would overthrow the officials if they could.

What does NJT, construed as a sufficient condition for legitimacy, have to say about the legitimacy of the two regimes? NJT implies that the Coercian regime is morally legitimate. Since (1) Coercian officials always enact the requirements of right reason and (2) subject judgments about the requirements of right reason are often mistaken, it follows that subjects are considerably more likely to do what right reason requires by following Coercian directives than by following their own judgments. Thus, the Coercian regime is morally legitimate under NJT.

In contrast, NJT does not indicate whether the Consentian regime is legitimate. It is true that the Consentian regime lacks the legitimating property described by NJT and is thus not legitimate *by virtue of* possessing that property; for the Consentian officials are no better than subjects to discern the requirements of right reason. But this simply tells us that Consentia does not satisfy the test of legitimacy described by NJT. This construction of NJT does not preclude there being other sufficient conditions for moral legitimacy. If there are, it might be that Consentia is legitimate by virtue of satisfying some other sufficient condition for legitimacy. Accordingly, NJT tells us that Coercia is legitimate, but says nothing about Consentia.

Both results are problematic. Consentian subjects consent to the authority of the officials and take the internal point of view towards the legal system. Though they sometimes flout the law, subjects generally regard the state with respect and frequently comply with laws they oppose out of respect for the regime. And it is not hard for us to see how subjects might reasonably arrive at such an attitude. Recognizing (1) that some kind of central body is needed to regulate behaviour and (2) that no citizen is more likely than the next to discern the requirements of right reason, citizens sensibly realize the value of a democratic system that provides them with ample opportunity to be heard. Indeed, it is not immediately clear how to construct a morally superior legal system out of these admittedly

artificial materials. Given the centrality of such features to much contemporary philosophical thinking about legal systems like that in the US, it is something of an embarrassment that a principle that purports to be a 'normal' justification of authority cannot tell us anything about the legitimacy of *Consentia*.

More worrisome, of course, is what NJT tells us about *Coercia*. In marked contrast to the situation in *Consentia*, *Coercian* subjects comply with the law sometimes out of fear of sanctions and sometimes out of respect for right reason, but never out of respect for the *Coercian* regime. Even when the laws simply require what *Coercian* subjects recognize as being morally obligatory, there is a sense in which they nonetheless regard the regime as purely coercive. By hypothesis, the subjects want badly to overthrow the officials, who are indifferent to subject resentment and continue to enforce their laws by means of coercive sanctions without providing any meaningful opportunity to be heard. Given the level of psychological resistance to the *Coercian* regime, it makes sense to characterize that regime as purely coercive because it is being imposed on the subjects against their will. This, coupled with the view of officials about their racial superiority, seems to disqualify the *Coercian* regime from being morally legitimate; more than mere impeccability is needed to entail legitimacy.

Accordingly, NJT is problematic as a sufficient condition insofar as it validates *Coercia* because *Coercia* seems to be a clear example of an illegitimate regime. Further, though a sufficient condition for legitimacy need not validate every legitimate regime, one would expect a sufficient condition purporting to be a *normal* justification to validate *Consentia*. NJT seems problematic as a normal justification, then, because it does not validate *Consentia*.

3. NJT is inconsistent with substantive constraints on legitimacy

It also seems reasonable to think there are substantive, content-based limits to the scope of legitimate lawmaking authority. While law may in some sense 'claim' authority to restrict freedom without limit, that claim is likely to strike modern sensibilities as an implausible claim. It is reasonable to think, for example, that the state may legitimately restrict expressive and sexual behaviour only up to a point. It seems clear, for example, that it would be wrong for the state to prohibit political speech or 'French kissing' between consenting adults.

Construed as a sufficient condition for legitimacy, however, NJT acknowledges no such limits. If NJT is a sufficient condition for legitimacy, the scope of the state's legitimate authority over the individual is unlimited. Indeed, a legitimate state may intrude into even the most intimate aspects of a person's life as long as it demands of its subjects what right reason requires. Suppose, for example, that right reason requires that people abstain from any non-procreative sexual behaviour whatsoever and from any sexual behaviour that does not take place within the context of a marriage that is consummated in a Roman Catholic ceremony (as would presumably be the case if the core doctrines of Roman Catholicism are true). Then any authority that is legitimate under NJT may justifiably

criminalize sex that occurs outside of such boundaries regardless of what citizens believe about it. Construed as a sufficient condition for legitimacy, NJT is incompatible with the existence of any protected spheres of citizen autonomy.

Arguably, a subject can waive any claim she might have to such protected spheres, as may impliedly have been done in *Consentia*—though each subject would have to do this for herself and presumably would have to retain some sort of entitlement to revoke that waiver. But nothing like that happened in *Coercia*. By hypothesis, the subjects in *Coercia* despise the Coercian officials and have not consented to their authority in any way. Yet the Coercian officials, being impeccable, require all and only behaviours required by right reason. For this reason, the regime in *Coercia* exhibits both of the sorts of deficiency discussed above: it respects neither procedural nor substantive limits on its lawmaking activities.

It is not entirely clear whether Raz would find such implications troubling. For example, Raz observes that the state typically claims its authority is unlimited in scope:

Governments typically claim authority to govern any and all aspects of their subjects' lives. They do in fact interfere with their education, their relations with their parents, with their children and spouses. They decide which sexual activities are allowed and when and where. They decide how many children one may have, how many spouses, and of what age and race. They decide whom one can visit and befriend, and where and how much, which religion one may follow and where, and in what form, which ideas one may propagate and by what means.³⁸

Unfortunately, there is, to my knowledge, no passage in which Raz addresses the issue of whether he believes the scope of legitimate authority extends to every requirement of right reason no matter what the topic.

But if right reason dictates some set of behaviours regarding any of the issues described in the above paragraph, then NJT, construed as a sufficient condition for legitimacy, would allow the state coercively to enforce the behaviours in that set. For this reason, this construction of NJT implies that the scope of a legitimate authority is substantively unlimited; there is no requirement whatsoever that subject autonomy *ever* be respected. And this seems to refute the claim that NJT operates as a sufficient condition for moral legitimacy.

There is a second problematic respect in which the scope of even legitimate authority is unlimited under NJT. Insofar as NJT justifies authority only in terms of its comparative superiority in discerning the requirements of right reason, it places no constraints on what that authority can do in the way of enforcing its directives. For how an authority enforces its directives does not logically bear on whether its directives are more likely than the subject's judgment to reflect the requirements of right reason. Thus, for example, if right reason

³⁸ GC 85.

requires that persons not park in front of fire hydrants, an authority otherwise legitimate under NJT may without compromising its legitimacy enforce that requirement by ridiculous means, including by means of the death penalty. Indeed, construed as a sufficient condition for legitimacy, a regime like Coercia may legitimately enforce all of its directives by means of the death penalty.

The problem here arises because NJT, by its own terms, governs only the ability of authority to discern what *authority-subjects* ought to do; apart from this, it does not define any substantive behavioural constraints on what authorities may permissibly do *qua* authorities. As long as the authority's substantive directives better reflect what right reason requires of *subjects* than the subjects' own judgments, NJT entails no constraints on authoritative acts. In particular, NJT entails no constraints on what sort of coercive measures may be used to enforce the directives of legitimate authority.

This renders NJT unique among mainstream theories of legitimacy. While it is possible to take a natural law position that has this result, there is nothing in even the most conservative natural law views that *logically precludes* there being constraints on what a legitimate state may do in the way of enforcing its directives.³⁹ For the claim that a state is legitimate only insofar as it conforms to natural law principles does not tell us anything about whether there is a natural law principle limiting what may be done in the way of enforcing morally legitimate directives. Accordingly, if NJT's failure to define limits on what directives can legitimately be enforced is not problematic, its failure to define limits on what can be done by way of enforcing a directive is.

(iv) *NJT as a necessary condition for moral legitimacy*

Theories that ground legitimacy in subject consent pose a problem for NJT construed as a necessary condition for moral legitimacy. Consent theories, which include social contract theories, disagree on the role that consent plays in the justification of authority. Classical consent theories assert that authority is legitimate to the extent that its subjects expressly or impliedly consent to it.⁴⁰ In contrast, John Rawls's consent theory asserts that authoritative directives are legitimate only to the extent they cohere with principles to which one would consent from an original position of ignorance.⁴¹ On all of these accounts, the consent of the subject—whether express, tacit, or hypothetical—is a sufficient condition for the legitimacy of state authority. If *any* of these theories are correct, then satisfaction of NJT is not necessary for moral legitimacy.

Raz finds consent theories problematic because they 'presuppose that consent is a valid source of obligations but [do] not explain why'. As Raz points out, consent is not always sufficient to create an obligation; consent to perform a

³⁹ See, e.g. William Blackstone, *Commentaries on the Law of England* (Chicago: The University of Chicago Press, 1979).

⁴⁰ See, e.g. John Locke, *Second Treatise of Government* (Indianapolis, IN: Hackett Publishing, 1980).

⁴¹ See John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).

violent act, for example, cannot give rise to a moral obligation to do so. And Raz believes that, given the extensive authority the state claims, a citizen's consent to the state, by itself, cannot obligate the citizen to obey state directives: '[I]t is hard to see what justification there can be for the binding force of consent to the authority of governments, given the extensive powers modern governments claim to have'.⁴²

It would be wrong, however, to think consent is utterly inert. There are clearly instances in which a subject can change existing duties and rights by giving consent. For example, my consent to allow you to enter my home transforms what would otherwise be impermissible trespass into a permissible act. Here my consent is sufficient to alter our moral relations. My consent waives my right that you not enter my home and thereby releases you from the corresponding obligation. And, for the most part, nothing more is needed to waive a right than the informed and voluntary consent of the right-holder.

But while it is clear that one can sometimes waive a right merely by consenting to do so, Raz is right to be sceptical that one can ever take on an obligation merely by consenting to it. Consider the case of promising. Suppose Tom wants to meet Mary at the library and says to her, 'Mary, I promise I will be at the library on Tuesday'. Mary responds, 'I don't believe you' and leaves the room with no intention of meeting him. Later on, she has a change of heart and goes to the library looking for Tom. Tom never shows up at the library because he believes there is no point in doing so. Though Tom probably should have gone anyway, it seems clear that he did not violate any obligations by not turning up there.

One might think that Tom's promise created an obligation from which Mary's subsequent behaviour released him, but this is different from a case where Mary accepts the promise by saying, 'I'll see you there' and subsequently releases Tom from it. In this latter case, Tom walks away knowing that Mary expects him to be there because her behaviour indicates she will rely on the promise. In the original case, however, Tom leaves with the impression that the situation between him and Mary is exactly what it was before he uttered 'I promise I will be at the library on Tuesday'. He rightly takes Mary's behaviour to mean that she gives no effect to those words and will not rely on them. And Tom's understanding here matches up with Mary's intentions; it is clear that what motivates her behaviour is not an intention to release Tom from any obligation inasmuch as she does not believe he can be trusted.

Tom's consensual commitment to undertake an obligation as expressed in his promise is, by itself, insufficient to create an obligation on his part to meet Mary. As counterintuitive as it may seem given the common understanding of a promise as a purely unilateral act, Mary also needs to do something for Tom's promise to create an obligation: for the promise to take, so to speak, Mary must accept it. While promises involving commitments to assume substantial burdens

⁴² GC 85–6.

may require more than acceptance by the promisee, it is reasonable to think that no promise can create an obligation unless it is accepted. Since the acceptance of the promisee is necessary to create an obligation, the unilateral commitment of the promisor alone is insufficient to create an obligation.

For somewhat different reasons, it also seems clear that subject consent to follow authority is not, by itself, enough to give rise to a moral obligation to obey authority. Insofar as the legitimacy of authority is wholly grounded in the subject's consent to it, its continuing legitimacy depends on her continuing consent, which can be withdrawn at any moment. Thus, for example, if my moral obligation to obey authoritative directives is grounded *entirely* in my consent, that obligation can be extinguished at any time by my withdrawal of consent. Indeed, on this account, I am no more bound by authoritative directives than by advice I am free to accept or reject at my discretion, because I am free to withdraw and renew my consent to authority at my discretion.

Contract theories, however, usually involve the idea that all persons in the relevant group have expressly or impliedly consented under conditions in which it is known that the others have all consented under similar terms.⁴³ If members of a group all consent (in the relevant sense) to abide by an authority's directives, then each member is giving up options otherwise available to her. If, in addition, each member's willingness to give up those options is conditioned upon the willingness of others in the group to do so, then each member commits to giving up these options in reliance upon the commitments of the other. This mutual reliance plays a necessary role in binding the members of the group to their commitments.

It is important here not to understate the role of mutual reliance and its implications. That mutual reliance among subjects is necessary to create obligations implies that a convergence of unilateral acceptances of authority, by itself, is not sufficient to create obligations. If the subjects are not explicitly aware of one another's acceptances and do not rely on them, then a convergence of acceptances is a matter of coincidence in the sense that each acceptance is independent of the others and hence gives rise to no legitimate expectations. A convergence of utterly independent acceptances is probably not sufficient to obligate each of the acceptors.

While the notion of mutual reliance does not entail that the acceptances are bargained for, it does entail that the parties know that the others have consented to the authority. To rely on someone else is to form an expectation about that person's behaviour and to take that expectation into account in deciding what to do. To rely on another person's acceptance is hence to form an expectation that the person will behave appropriately given her acceptance of authority. But such reliance is not possible without awareness that the other person has also consented. Accordingly, we can characterize the relevant cluster of relations giving rise to

⁴³ Theories involving something like hypothetical consent require a different story—one that is not relevant here.

obligations in such circumstances as a joint commitment among the subjects. A commitment that consists of mutual consent and reliance among the subjects, then, appears sufficient to create a moral obligation to obey.

In response, Raz would argue that '[a]greement or consent to accept authority is binding, for the most part, only if conditions rather like those of the normal justification thesis obtain'.⁴⁴ There are, of course, limits on the extent to which consent gives rise to moral obligations—even if that consent is bargained for or relied upon by other people. As Raz points out, consent to regard a directive as authoritative implies certain restrictions on how an authority determines which directives to issue. Thus, for example, a judge's decision that is based on the outcome of a coin-flip falls outside the scope of the judge's authority. Likewise, mutual consent and reliance is not enough to rescue a bargain if the bargain is sufficiently unfair to one of the parties—either because of the bargain's content or because it was not negotiated at arms length.

But these are exceptional circumstances and not the general rule with respect to the relation between consent and authority. If (1) the parties are capable of giving effective consent to authority, (2) the parties consent to the authority *and* to some kind of enforcement mechanism, (3) the consent is secured in a fair manner, (4) each party gives up something of value in relying on the other, (5) the authority's directives are based only on considerations of what right reason requires, and (6) the authority decides in good faith what ought to be done according to right reason, then it seems reasonable to think that the directives of the authority give rise to moral obligations and may hence be permissibly enforced against the subjects by the state. Indeed, if one consenting party foreseeably relies to her detriment on the consent of the other to abide by a directive, it seems unjust not to enforce the directive against the latter.

Notably, several of these conditions operate as constraints on the content of authoritative directives. Just as there are limits to what consent can obligate a subject to do, there are also limits on what a consensual authority can obligate a subject to do. Under no circumstances, for example, could an authority morally obligate a subject to kill a person for the reason she has offended the authority. Conditions (1), (3), (5), and (6) operate to define these substantive limits. Condition (5), which resembles the Dependence Thesis, ensures that the authority's directives reflect *only* those reasons that *objectively* apply to the subjects. A competent party whose consent is secured in a fair manner—conditions (1) and (3)—presumably consents only to authorities whose directives fall within a range of prudentially and morally acceptable alternatives. Similarly, conditions (1) and (3) imply that the subjects consent only to authority that is competent to assess the objective reasons. Condition (6) operates to ensure that the authority makes decisions sincerely and to the best of her ability to decide what ought to be done according to right reason.

⁴⁴ ALM 214.

While these conditions clearly limit the discretion of authority in deciding what ought to be done, they fall well short of requiring that the authority do a better job of assessing what right reason requires than the subjects. The condition implying that authority be competent ensures there will be instances in which some subjects will more likely comply with right reason by following the authority's directives than by following their own judgments of what right reason requires. But competence is not infallibility; hence there will also be instances in which the subjects would do better by following their own judgment. Indeed, and this is the important point, conditions (1) through (6) do not logically preclude there being an authority whose legitimacy extends even to consenting subjects who, *as a general matter*, are more likely to comply with the demands of right reason by following their own judgments than by following the authority's directives. Thus, if satisfaction of (1) through (6) is sufficient for legitimacy, then satisfaction of NJT is not *necessary* for consent to authority to give rise to a moral obligation to obey the directives of that authority.

It is, of course, highly unlikely that there are any legal systems in which conditions (1) through (6) are satisfied; certainly this is not true of any existing legal system. But the point here is not to articulate a general theory of legitimacy that would systematize intuitions about the legitimacy of any particular existing state. I do not and need not claim, for example, that (1) through (6) are necessary conditions for the legitimacy of a political authority. All that is being claimed here about conditions (1) through (6) is that *if* these conditions are satisfied, the directives of a legitimate authority give rise to enforceable obligations. And if this is correct, as should seem eminently plausible, then NJT does not state necessary conditions for moral legitimacy.