

Liberalism and Individual Positive Freedom

Author(s): John Christman

Source: *Ethics*, Vol. 101, No. 2 (Jan., 1991), pp. 343-359

Published by: The University of Chicago Press

Stable URL: <https://www.jstor.org/stable/2381867>

Accessed: 09-05-2020 17:45 UTC

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

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Ethics*

Liberalism and Individual Positive Freedom*

John Christman

Among the various principles central to the tradition of liberalism is the claim that the state should protect individual liberty without dictating the goals and purposes espoused by free people. This tenet has been given expression by the general liberal emphasis on negative liberty—the claim that the liberty of a person is strictly a function of the restraints that the agent faces in the carrying out of her decisions (however the concept of a restraint is construed). The person—the complex set of functioning capacities and the forces that condition them—is not to be counted in the calculation of the freedom of that agent. It is a constant, so to speak.

However, purveyors of the notion of *positive* liberty insist that the person and her capacity to formulate her desires, values, and goals is a crucial element in the calculation of the freedom of the agent. However, these writers have not responded to doubts from traditional liberals who are famously critical of the inclusion of any positive components in the concept of freedom. In this article I wish to respond by claiming that a certain notion of positive liberty can be defended against the “classic liberal” objections to the notion.

The most challenging of these criticisms can be culled from Isaiah Berlin’s “Two Concepts of Liberty” and remain forceful even after much discussion of them.¹ These objections are that the concept of positive

* Earlier versions of this paper were read at the University of California, San Diego (December 1987) and at the American Philosophical Association (APA) Central Division Meetings (April 1989) in Chicago. I would like to thank the members of those audiences and, in particular, Richard Arneson who was the commentator for the APA presentation. I am also grateful to the editors of *Ethics* for their helpful suggestions.

1. Isaiah Berlin, “Two Concepts of Liberty,” in his *Four Essays on Liberty* (Oxford: Oxford University Press, 1969), pp. 118–72. See also Isaiah Berlin, “From Hope and Fear Set Free,” in his *Concepts and Categories* (New York: Viking, 1979), pp. 173–98. Much work has been done on these questions in the way I am concerned with here: see, e.g., Richard Arneson, “Freedom and Desire,” *Canadian Journal of Philosophy* 15 (1985): 425–48; Charles Taylor, “What Is Wrong with Negative Liberty,” in *The Idea of Freedom*, ed. Alan Ryan (Oxford: Oxford University Press, 1979), pp. 175–93; John Gray, “On Negative and Positive Liberty,” in *Conceptions of Liberty in Political Philosophy*, ed. Zbigniew Pelczynski and John Gray (New York: St. Martin’s, 1984), pp. 321–48. Those who have recently defended

Ethics 101 (January 1991): 343–359

© 1991 by The University of Chicago. All rights reserved. 0014-1704/91/0102-0005\$01.00

liberty is a paradoxical (if not incoherent) notion and that the promotion of positive liberty is inconsistent with the most basic principles of a free (liberal) society. In responding to these worries, spelled out in what I take to be a new (or newly formulated) way, I wish to galvanize and clarify a notion of positive liberty that is both faithful to the concerns of the tradition in which it is central and that captures some of the basic ideas of liberal principles of justice. In particular, I will defend the plausibility of the concept of *individual* positive liberty, a notion which I think is immune to those central liberal objections.

THE CONCEPT

The idea of positive freedom has been variously rendered. The notion I will be defending represents an attempt to capture the requirement that free agents must be, in a fundamental sense, self-governing. This notion will be equivalent to the concept of individual autonomy as that is sometimes discussed. The distinction between this idea and the concept of *negative* liberty is between seeing freedom as simply the absence of restraints (of any kind) and seeing freedom as (in addition to this) the capacity for self-mastery and self-government.² Defenders of positive liberty claim that the concept of a restraint—no matter how embellished an account one gives of it—will not capture the ways that people can be manipulated and conditioned in relation to the very makeup of desires and values.

Imagine, for example, a woman who is raised in a culture which fiercely inculcates in her the idea that women should never aspire to be anything but subservient and humble domestic companions to their husbands, no matter how unhappy this makes them or how abusive their husbands are. Imagine further that this person is suddenly placed in a new culture where opportunities abound for women to pursue independent

theories of freedom that contain positive conditions include Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon, 1986); S. I. Benn, *A Theory of Freedom* (Cambridge: Cambridge University Press, 1988); and Carol C. Gould, *Rethinking Democracy* (Cambridge: Cambridge University Press, 1988), chap. 1. I do not see, however, that any of this recent work has succeeded in both developing a detailed account of positive liberty and responding adequately to Berlin's challenges.

2. Gerald MacCallum argued, of course, that there is not a true distinction between positive and negative liberty ("Negative and Positive Freedom," *Philosophical Review* 76 [1967]: 312–34). His view is that freedom of any sort is always a triadic relation among a person, an action, and a restraint. But even accepting MacCallum's formal elements, it is nonetheless a crucial aspect of classical liberal doctrine to regard freedom as being increased or decreased *simply* as a function of the presence or absence of restraints. John Rawls, e.g., discusses the MacCallum variables and admits that there are possible variations in what shall be counted to fill in for the 'X' placeholder (the concept of the person). But all he considers for such variants are the "various kinds of agents who may be free—persons, associations, states" (*A Theory of Justice* [Cambridge, Mass.: Harvard University Press, 1971], p. 202). So if one narrows this range to persons, then on the liberal view, the only variable left in the quantification of freedom is the number of restraints.

activities. She nevertheless shuns these opportunities and remains married to an oppressive husband from the old culture. The only "restraint" she faces (to pursuing the opportunities for an independent life-style) are her desires themselves (which remain the sort she was taught to have). She simply does not *wish* to act in any other way, turning a deaf ear to the reasons people give her to consider a less subservient posture. (Imagine that her husband abuses her but tells her she can leave him any time she wants, and she continues to want to stay.) Defenders of positive liberty insist that such a woman is unfree, and precisely because the processes by which her character and values were developed were themselves oppressive. These methods did not allow her to reflect on her emerging values in light of reasonable alternatives. So the presence of opportunities—the absence of restraints—is irrelevant to the true nature of her unfreedom. This shows that negative liberty is simply incomplete as a full accounting of human freedom. The free person must be guided by values that are her own. This is what the idea of positive liberty attempts to express.

Positive freedom, in the sense I want to discuss, will apply principally to *individuals*. I should therefore say a word about the connection which has traditionally been recognized between liberty in the positive sense and political participation—self-government in the collective sense. The relation between positive freedom and political participation is a complex one, but it is a connection which is not, on the view discussed here, *conceptually* necessary. The connection traditionally recognized can be explained this way: self-government means being guided by forces which are self imposed; the institutions of one's government and society to a large extent shape a person's attitudes and values (as well as actions); so unless one participates in the fair democratic institutions of government, it cannot be said that those laws and social forces truly emanate from one's will; and hence, only via participation is one self-governing. This argument depends heavily on the premise concerning the determining force of social and legal institutions. To maintain the conceptual separateness of the notion of positive liberty and democratic participation, one need only point out the contingent nature of the linking premise that our attitudes and values are molded by our society (in some strong sense). In a modern (and large) industrial society, a good many of my concerns are not severely dictated to me by the reigning governmental institutions of the day. So insofar as this is true, I can be to a large extent (individually) self-governing even if the institutions of the state and I keep a respectful distance. But in any case my claim here is only that the notion of *individual* positive liberty is of a piece with the tradition and also does not make participation in democratic institutions a conceptual necessity.

For an individual to be self-governing it at least must be the case that she is not moved by desires and values that have been oppressively imposed upon her, even if she faces no restraints in performing actions

such desires motivate. Her character must be formed in a certain manner. What is needed, then, is an account of how desire changes take place, which is an expression of the ideal of the fully free person. Preference changes cannot be the result of oppressive conditions or blind, unreflective conformity to limited choices. Self-mastery means more than having a certain attitude toward one's desires at a time. It means in addition that one's values were formed in a manner or by a process that one had (or could have had) something to say about. It is in this way that positive freedom will be a property of the "true self," but this self need not be metaphysically set apart (e.g., from the "phenomenal" self) or ontologically mysterious.³

There has been much work in recent years on the concept of individual autonomy which is relevant here. Typical of this is the view that "a person is autonomous if he identifies with his desires, goals, and values, and such identification is not influenced in ways which make the process of identification in some way alien to the individual."⁴ This approach has faced various objections, most of which focus on the vagueness of the identification requirement and the threat of an infinite regress of the conditions.⁵ As my remarks so far have suggested, I would urge that what is needed is an account, at the level of preferences, of what processes of self-change preserve autonomy and which ones do not, an account which does not depend on the condition of identification or on the requirement of an infinite string of self-chosen desires. This can be accomplished, I think, when it is insisted that the conditions of autonomy essentially bear on the *formation* of preferences, not on their structure at any one time. The account would go something like this: whatever forces or factors explain the generation of changes in a person's preference set, these factors must be ones that the agent was in a position to reflect upon and resist for the changes to have manifested the agent's autonomy. In addition, this reflection and possible resistance cannot have been the result of other factors which—as a matter of psychological fact—constrain self-reflection.⁶

This latter condition is needed to prevent an infinite regress of self-chosen desires as being necessary for autonomy. What must be true of the agent's acceptance (or rejection) of the processes of preference change

3. For defense of this last claim, see John Christman, "Autonomy: A Defense of the Split-Level Self," *Southern Journal of Philosophy* 25 (1987): 281–94.

4. Gerald Dworkin, "The Concept of Autonomy," in *The Inner Citadel: Essays on Individual Autonomy*, ed. John Christman (New York: Oxford University Press, 1989), p. 61. Dworkin has revised this view in *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988), chap. 1.

5. For a discussion of these criticisms, see John Christman "Constructing the Inner Citadel," *Ethics* 99 (1988): 109–24.

6. The account I discuss here is more fully explained and defended in my "Autonomy and Personal History," *Canadian Journal of Philosophy*, vol. 20 (1990). In that paper I add a condition that the agent cannot be "self-deceived" to be autonomous. This, as I explain there, is an elaboration of the rationality condition I discuss below.

is that at some level this was done in a “clear-headed” manner. That is, such things as drugs or emotional stress—that is, any factors which we know cloud a person’s normal ability to reflect at all—were absent in the self-reflective processes that took place (or would have taken place). This is not to say that these factors must be absent at every level, for someone might autonomously choose to expose herself to reflection-inhibiting factors; but then in the case of *that* choice, or one at some more basic level, reflection-inhibiting factors must be absent. This requirement avoids a regress of the conditions for freedom.⁷ It also captures the intuition that freedom demands more than the condition that desires must be self-chosen, it must also be the case that these (meta-) choices are made under conditions free of external manipulation and interference.

We can articulate, then, the following conditions for autonomous generation of preferences. A person P is autonomous relative to some desire D if:

1. P was in a position to reflect upon the processes involved in the development of D;
2. P did not resist the development of D when attending to this process of development, or P *would not have* resisted that development had P attended to the process;
3. The lack of resistance to the development of D did not take place (or would not have) under the influence of factors that inhibit self-reflection (unless exposure to such factors was autonomously chosen, in which case that choice had to be made without such factors); and
4. The judgments involved in this self-reflection, plus the desire set that results, are minimally rational for P.⁸

Of course much needs to be explained in these conditions. A person “attends to” the development of a desire when she is in a position to focus on the processes and conditions that led to the adoption of that desire. That is, a relevantly full description of the steps of reasoning or the causal processes that led her to have this desire is available for her possible consideration. This reflectiveness assumes that an agent can become aware of the beliefs and desires that move her to act. Call this the “transparency” of her motivating reasons. What I mean by this is the ability of an agent to bring to conscious awareness a belief or desire—either in the form of a mental representation or a proposition—and concentrate on its meaning.⁹

7. The regress is avoided since all that is required at this level is that the sorts of factors I describe are absent, not that a further level of choice is necessary.

8. This account is similar in some respects to Gerald Dworkin’s revised account (*The Theory and Practice of Autonomy*, chap. 1). My view, however, differs in some crucial details, e.g., in the addition of conditions 3 and 4 as well as my focus on preference *formation*.

9. Herbert Fingarette, in *Self Deception* (London: Routledge & Kegan Paul, 1969), describes the act of “spelling out” to oneself one’s beliefs. This is similar to what I have in mind here. This notion of transparency might be resisted by those influenced by Freud

The “processes” that give rise to a change in desire (or the development of a new desire) are various and can be described at any number of levels. On the one hand, the process can involve a change in a belief—the agent comes upon new information—where she can, in principle, examine the reasons for the new belief and its relation to other things already desired.¹⁰ In these cases, the process—a reasoning process—must be one the agent is guided by without irrationality. Preference changes must be the result of deliberations (perhaps hypothetical ones the agent would have if she turned her attention to the question) that do not involve inconsistencies and, by implication, mistakes in logical inference. I will return to this requirement presently. In the other kinds of desire formation processes, no alteration in the set of beliefs of the agent takes place. The explanations of such preference changes will be straightforwardly causal. The reason I have a desire to sleep right now, for example, is that I didn’t sleep enough last night and the coffee I drank has worn off (plus a variety of physical facts about me). No epistemic element is relevant to the explanation of this new desire. In these cases it is hard to be specific about the level of description of these processes that the agent must consider, for theories vary as to the exact causal explanation of these transitions. But these issues do not need resolution for our purposes. All that must be true is that the agent would not resist, that is, be moved to try to counteract, the process, were she to understand it. The level of description, then, can be any one which the agent is capable of considering and which is not patently inconsistent with some verifiably true description at another (deeper) level. For the most part, though, causal processes that do not involve epistemic steps in the development of a desire will not be ones that the agent could resist even if made aware of them. No amount of detailed information about my bodily mechanisms will make it possible for me to stop being tired right now. Hence, this is not an autonomous desire since I would prefer to go on working longer without this nagging urge to sleep.

The motivating idea behind this theory is that autonomy is achieved when an agent is in a position to be aware of the changes and development of her character and of why these changes come about. A qualification is in order however: I am assuming here that this model of autonomy applies to adults whose childhoods have not been manipulative and autonomy-inhibiting. Admittedly, this is a highly artificial assumption. However, the model described here can plausibly be used to evaluate

who are convinced that much of our motivational structure is not immediately transparent to us (without therapy, dream interpretation, or the like). I do not wish to dispute this here. I only claim that *insofar as* a person’s motives are subject to reflective consideration by the agent under normal conditions, she is autonomous. If therapy is necessary to make this possible then therapy is necessary to make autonomy possible.

10. Jon Elster, in *Sour Grapes* (Cambridge: Cambridge University Press, 1983), provides a topology of changes in beliefs and desires that is informative here.

child-rearing techniques concerning their tendencies to promote autonomy. I assume that we can evaluate the factors by virtue of which we developed the preferences we did *apart* from the particular preferences these processes produced. If I look back on my past and see that much of my character was formed by educational and parental practices that I would not want to have been molded by, or if my conditioning has been so manipulative that I cannot even reflect clearly on these events and processes at all, then I am not autonomous relative to those aspects of my character. As I do not have this attitude toward the development of at least some of my most basic personality traits, I regard them as autonomous aspects of my present self. This is so even if, at the time, I had nothing to say about them.¹¹

The fourth condition in this model must now be discussed. The tradition of positive liberty is rife with claims connecting “true” freedom with the demands of *reason*.¹² But to what extent must the judgments involved in the processes of self-reflection demanded by positive freedom be *rational*, or in what *sense* must they be rational?

Criteria for rationality vary, ranging from the demand for consistency of beliefs and desires, to requiring the choice of the most effective means for one’s ends, to having “good evidence” for the beliefs upon which one’s desires depend, etc.¹³ And it is indeterminate which of these views is the most plausible necessary condition for positive liberty. Notice, however, that this range of demands for rationality can be separated into what can be called “internalist” or “subjective” accounts of rationality and “externalist” or “objective” accounts.¹⁴ On an internalist account, the property by which an action is considered rational for an agent bears only on those beliefs and desires actually “internal” to the agent, not on the relation between those beliefs and the world (i.e., a relation of fit or

11. Some have read this model as implying that an autonomous agent cannot, ex ante, disapprove of some desire that she might develop. This would be implausible since we often would not want to have certain preferences until we *do* have them (e.g., being in love with a certain person). But the condition of reflection I am putting forth here concerns principally judgments about the *processes* of preference changes, not the results. What matters for my autonomy is my attitude about how I came to *be* in love, not whom I end up loving. (I am grateful to Richard Arneson for aid in clarifying this point.)

12. See Berlin, “Two Concepts of Liberty,” pp. 145–54.

13. I will here be assuming a distinction between “conditional” and “brute” desires. The latter are simple desires which do not depend on the truth of some belief the agent has (the desire for pleasure is an example). Conditional desires are ones an agent has which are contingent on the truth of such beliefs, such as the belief that getting the thing desired will enable the agent to achieve some further aim. Some writers have also discussed what have been called “value attributing” desires. These are nonconditional desires which involve beliefs about the object of the desire, for they also involve judgments of the *value* of the thing. I will consider this third kind of desire as basically conditional, for the issue of whether beliefs upon which desires depend are justified will apply to value attributing desires in the same way as it does to conditional preferences.

14. Similar to this is Richard Brandt’s distinction between “subjective” and “objective” rationality (*A Theory of the Good and the Right* [Oxford: Clarendon, 1979], pp. 72 ff.).

accuracy). Usually what is demanded is that the beliefs (upon which the person's conditional desires are based) are consistent and the desires (whether conditional or "brute") are transitive.¹⁵ This can be contrasted with externalist criteria, whereby an agent is rational only if she has gathered (objectively) adequate evidence to justify her beliefs (upon which the desires she entertains rest). On this account, lacking relevant information upon which a desire is founded renders that desire irrational.¹⁶ The most stringent version of an external rationality condition, the one which is the locus of the criticism discussed below, is one which requires that the agent conform her desires to the correct *values* as well as facts. One way to capture the distinction between internalist and externalist conceptions of rationality is this: the internalist would only demand that a person act for *reasons* (perhaps ones which meet some requirement of consistency), while the externalist demands that the free agent must act in accordance with *reason*, where that includes knowledge of the truth, both about the world as well as morality.

On the model of positive liberty being defended here, a minimal, internalist rationality requirement is added to the procedural conditions for the development of the desires of the autonomous self. This means that agents who are acting on the basis of inconsistent beliefs or intransitive desires are not acting freely. This condition must be subject to some principle of charity since likely no one has completely consistent beliefs and transitive desires if all the logical implications of both are considered. What the requirement demands is that no *manifest* inconsistencies figure in the desire in question. However, there is no requirement that the beliefs in question (upon which conditional desires rest) be reliable, that they fit the facts. Similarly (and perhaps more familiarly), the brute desires of the agent cannot be appraised on the basis of their rationality.

15. The conception of rationality as the maximization of expected utility, e.g., is internalist as long as there can uncontroversially be attributed to the agent an overall desire to maximize happiness. In that case, the model demands only that desires be transitive (with the desire to maximize utility overriding all others) and that beliefs—about relevant probabilities and the commutations—be consistent (i.e., that the laws of probability are obeyed). This is internalist because it contains no stipulations concerning whether the probability beliefs that are utilized be justified (externally). Furthermore, we can assume that all other requirements for rationality are met, e.g., that the actions are caused by the desires and beliefs "in the right way" (cf. Donald Davidson, "Rational Animals," *Dialectica* 36 [1982]: 317–27). Also, by 'consistency of beliefs' I mean that the set of beliefs could all be true in a single possible world. In the case of consistency of preferences, it is common to require that they be transitive, complete, and continuous. These are very stringent requirements, though, as most people have not compiled a complete ranking of all the available objects of preference. So by 'consistency of desires' I will mean simple transitivity of those desires plus consistency of the beliefs upon which they rest (if any).

16. For example, Robert Audi demands that the beliefs upon which the desires of a rational agent rest must be "well grounded"—that they be reliable in an externalist epistemic sense ("Rationality and Valuation," in *Social Action*, ed. G. Seebass and R. Tuomela [Boston: Reidel, 1985], pp. 243–78). For a general discussion of the relation between rationality and autonomy, see Richard Lindley, *Autonomy* (New York: Macmillan, 1986), pp. 21–70.

Also, including an internalist rationality requirement for freedom does help capture a traditional claim that defenders of positive freedom want to make, that actions which emanate from *uncontrollable* desires—impulses and compulsions—are not free. But this conclusion need not rest on the traditional reasons relied upon: that “lower,” animalistic desires should be regulated or expunged by the “higher” self for freedom to be attained. On the present view, it is the manifest *conflict* of desires, at any level, which yields the judgment that the agent does not act freely.

This marks an improvement over traditional accounts since not all actions that result from “uncontrollable” impulses should plausibly be counted as unfree: a sprinter waiting in the starting blocks builds up a degree of energy and concentration that, when the gun is fired to start the race, she cannot help but lunge forward and start running; but we would not call such action unfree merely because (at the point just after the shot started the race) an irresistible desire figured in the ensuing activity. What *would* make the resulting act unfree is if the agent had other desires to stop which were overcome by the “force” of the one to run. So while compulsive desires themselves do not undercut freedom, preference sets that are in conflict do render the actions which emanate from them unfree.

Demanding an external rationality condition, however, is a different story. This amounts to the claim that all beliefs that figure in the actions of agents must be justified by available evidence, that they fit the facts. Below I will suggest why an externalist rationality requirement—either an epistemic or a value requirement—should not be added to the minimal internal condition just discussed as a condition of positive liberty. I conclude then that the self-reflection that autonomy necessitates must not involve manifest inconsistencies which bring into question the unity of the self. Hence, what is required of liberty is (only) minimal, internalist rationality.

I will now defend this model against the liberal objections that have been directed at notions of positive liberty.

THE LIBERAL OBJECTIONS

The liberal emphasis on the value of negative liberty motivates two central objections to its positive counterpart. Isaiah Berlin, still the clearest source of these arguments, urges that, on the one hand, positive liberty is paradoxical and commits one to counterintuitive conclusions about the ways that freedom can be increased, and, on the other, its promotion conflicts with other values to be protected in any just society (in particular, it allows the conceptual possibility that tyranny of individuals can be perpetrated in the name of freedom). I will call these arguments “the inner citadel” argument and “the tyranny argument,” respectively.

Berlin’s objections are motivated by the idea that if positive liberty is understood in the way it traditionally has been, the state is able to claim that it is acting for the greater freedom of its citizens when it

interferes with their character development and life plans.¹⁷ This way of framing the problem does collapse two very different issues: on the one hand, what is a plausible account of the concept of liberty, and, on the other, what sorts of infringements of a person's negative liberty (for the purpose of shaping her character) are justified. But both Berlin and I see that these issues are intimately connected, for insofar as the promotion and protection of social freedom is a foundational principle guiding the policies of the state, what one turns out to mean by "freedom" will in part determine what kinds of interventions into a person's development and life activities are justified. It should be understood in my discussion that the concept of autonomy I defend refers to something of *value*, and the principles of the just state should reflect this value. This implies that the exact characterization of this value will have implications for other principles that guide state policies (especially ones bearing on legitimate intervention).

The Inner Citadel Argument

This argument is roughly as follows: if liberty is construed as rational self-mastery, then I am made more free when, instead of removing restraints faced by my real wishes, I am manipulated into giving up those wishes. If freedom means doing what one wants, a person is made more free either by the removal of restraints on her choices or by the dissolution of the restrained choices themselves. This well-known difficulty is underscored this way by Berlin: "It is as if I were to say: 'I have a wounded leg. There are two methods of freeing myself from pain. One is to heal the wound. . . . But . . . [the other is to] get rid of the wound by cutting off my leg.'"¹⁸ I am made more free by retreating into "the inner citadel" of those core desires that form my true—free—self. Berlin labels this implication a "paradox" and suggests that any conception of liberty that entails it is thereby implausible.

Consider, for example, the familiar comparison between two types of slave: one is a person who hates her chains and longs for the things slavery prevents and is consequently miserable; the other is, like Epictetus, properly "adjusted" to her confinement and has expunged any of those desires that her situation has made impossible for her. Now on any conception of freedom which, like positive freedom, insists that an agent is free insofar as she can carry out those desires that are truly her own, then it must be concluded—implausibly—that the second slave is freer than the first.

Now as many have noted (and Berlin admitted), this example is not only a problem for the positive account of freedom. (In fact, I would

17. This, of course, is not Berlin's only line of argument in "Two Concepts of Liberty." Among other things, he is concerned to show how the concept of positive freedom *changed* crucially in the intellectual history of the West in the past two centuries. I am grateful to an editor of *Ethics* for calling my attention to the points to follow in this paragraph.

18. Berlin, "Two Concepts of Liberty," p. 135.

suggest that the positive account is considerably better equipped than the negative view to capture our intuitions concerning such happy slave examples.) On *any* account of freedom, the very conception of a restraint will need to make reference to actual or possible desires of an agent.¹⁹ If freedom consists in unrestrained *possible* desires, then the concept of liberty becomes vacuous due to the impossibility of enumerating restraints. For example, the books on my shelf apparently are not a restraint. However, if I decide to walk in a line that crosses through where they are (say a fire starts and they block what becomes my only escape route), then they are. This shows how the number of restraints I face at any given time is virtually immeasurable and hence so is freedom.

On the other hand, if objects are counted as restraints only if they frustrate *actual* desires, the status of objects as possible constraints changes according to alterations in desires of the agent. Thus any conception of liberty that counts restraints in accordance with desires faces the possibility that freedom is increased or decreased by changes in desires rather than in external circumstances. This points exactly to the paradox of the happy slave, or so it is claimed.

I think, however, that commentators on this problem have mislocated the locus of the paradox in such “happy slave” cases. What gives the examples the air of paradox is not simply the structural form it takes: “the presence of a desire and a restraint plus the removal of the desire equals an increase in freedom.” For certainly if I undergo a self-conscious and (let us say) rational program of character change, and, well after I am successful, some object or force appears that would have prevented me from fulfilling some previously discarded desire, then this object is no barrier to me, and it has no effect on my freedom. It is just like the books on my shelf (without the fire). And this scenario has the same “structure” as the happy slave cases. The paradox, on the other hand, arises when there is a suspicion that the preference change resulted from the very *presence* of the new restraint, bearing down on the agent and (forcefully) causing the change in desire. Calling *these* types of desire changes an increase in freedom is indeed implausible. What this shows is that, while it might be admitted that freedom can be a function of desires (since restraints are defined with reference to them), it cannot be that freedom is increased when desire changes take place directly *because* of the forceful presence of the new barrier.²⁰

The application to this problem of the model of positive liberty we have developed is direct: if a person acts upon desires that were not developed in accordance with the conditions set out above, then the person is not acting freely. Restraints can be characterized as those barriers to the carrying out of “autonomously formed” desires. Hence, if the

19. This is what Richard Arneson calls “the desire thesis” in “Freedom and Desire.” The admission by Berlin occurs on p. xx of *Four Essays on Liberty*.

20. Elster, chap. 3.

“happy slave” has expunged her desires for freedom only as a result of the oppressive presence of the restraints she faces, then she is not more free after the change. For it is not the case that her desires were formulated in a manner that she could have resisted (and, we imagine, she would have). The chains she feels still constrain those desires for free movement which were (we can suppose) freely formed according to the above formula. So they are still restraints, and she is not more free after the change.

But if desire changes take place, and these changes are autonomous in the above sense, the person will remain free (positively) when forces are introduced that hinder those jettisoned desires. If the desire change in question occurs autonomously then the presence of such a “barrier” does not affect the freedom of the agent. And this is a conclusion free, I think, of paradox and incoherence.

Richard Arneson doesn’t agree that this kind of move solves the inner citadel problem. On his view, even if desires are expunged by a process of self-reflective character change, the contented slave is not more free than the miserable one.²¹ I think, however, that his doubts stem from a failure to take seriously the results of the person’s change in character. When the “barriers” in question are things like locks and chains, and the actions being prevented seem so fundamental to normal human flourishing, it is hard to accept that a person could be *truly* indifferent to the presence of those restrictions. And this speculation is quite plausible. But it remains the case that if the Epictetan slave truly *does* extinguish the desires in question, in a way that is admittedly hard to imagine but not impossible, then those “restrictions” can no more be counted as restraints than can my books on a normal day. To make the example more believable, imagine that the desire changes take place well *before* the placing of the “restrictions,” so that there is no hint of the possibility that the newly placed objects are the (illicit) cause of the preference change. Imagine that the “happy slave” is a Tibetan monk who has spent the last several years in the same room meditating and sitting quietly (being fed by acolytes) and from which he will never desire to move. If chains are then put on the door to the room, a room he does not want to leave, then his freedom of action is simply unaffected by these chains.

So if a full account of desire formation along the lines described in the first section of this article can be worked out and defended, the idea of positive liberty which contains this as a component will no longer be subject to the inner citadel argument. And I should stress that this was a problem for both positive and negative conceptions of freedom, and it is the positive conception that is most able to accommodate it.

The Tyranny Argument

The second liberal objection to positive liberty—the tyranny argument—is this: according to the conception of positive freedom couched as rational

21. Arneson, p. 433.

self-mastery, rule of one's desires and actions by one's "true, that is rational, self" is the truest expression of freedom. But, "once I take this view," Berlin argues, "I am in a position to ignore the actual wishes of men or societies, to bully, oppress, torture them in the name . . . of their 'real' selves, in the secure knowledge that whatever is the true goal of man (happiness, the performances of duty . . .) must be identical with his freedom—the free choice of his 'true,' albeit often submerged and inarticulate, self."²² The claim is that seeing freedom as rational self-mastery allows the direct opposition between freedom in this sense and other (negative) liberties considered canonical among the principles of a just society. Moreover, it is counterintuitive to say, as defenders of positive freedom seemingly must, that such tyrannical intrusions into a person's life or activities or even thought processes can be taken as consistent with her liberty just because the justification for doing so is that the imposed desires or actions are in greater conformity with reason than are the agent's own. This, it is claimed, is both paradoxical and morally pernicious.

In order to respond to this argument it will be necessary to return to the issue of the exact conditions of *rational* self-mastery. Now if, as we have argued, the extent of the rationality requirement for positive liberty was the internal sense of rationality, then it is clear that the tyranny objection is avoided. For no second party (much less a tyrant in a position of power affecting many individuals) will be in the epistemic position necessary to justify intrusions on the basis of failed rationality of this sort. For it would have to be the case that the intruder knows more than the agent about the internal structure of her set of desires and beliefs and judges them to be inconsistent in some way. The practical impossibility of this scenario undercuts the force of this difficulty.

But what about a version of the more stringent, external, conditions of rationality? Requiring an external *evidence* condition for positive liberty allows for the tyranny objection only in a narrow range of cases: those where the "tyrant" has access to more factual information than the agent and thus can interfere with the (less informed) actions of the agent without thereby decreasing her freedom. But is this bothersome? Certainly, as before, the range of cases will be minimal: where, for instance, the information is indisputable, the agent had reasonable access to it, etc. On such views, to act unwittingly is not to act freely. And if I interfere with your unwitting actions I do not disrupt your self-government in any meaningful way. Most writers in the liberal tradition accept this as neither paradoxical nor pernicious.

This question, however, puts us face to face with the question of justified paternalism, the limit to which is, for many, respect for the self-government (autonomy) of the agent. But even J. S. Mill allowed for paternalistic intervention when, for example, a person is about to un-

22. Berlin, "Two Concepts of Liberty," p. 135.

wittingly step onto a faulty bridge. In fact most liberal writers on paternalism, who regard respect for the agent's autonomy as the limiting factor in justifying interferences for a person's own good, have no trouble with justifying interventions based on superior knowledge as consistent with the agent's autonomy. I take this as evidence that if positive liberty allows for these sorts of interferences "in its name," then this does not give rise, among liberals, to grounds for its rejection.

There is one reservation, though, that I have about including an evidence requirement among the conditions for freedom. It is that this renders the notion of free or autonomous agents essentially *indeterminate* for all those actions which are motivated by beliefs in empirical propositions. Although there are well-accepted theories of the thresholds which evidence gathering must reach for beliefs to be justified, it remains that these are mere thresholds. An action based on an alternative belief which rests on *better* evidence will be *freer*, and it will remain hazy (arbitrary?) where to draw the line demarcating free (informed) action from unfree (less informed) actions. I take this reservation, however, as merely cautionary at this point.

With all this ground cleared away, we arrive, I think, at the heart of the matter: that insofar as positive liberty requires an external value condition (in its demand of rationality), it is not in conflict with the severest form of tyranny—interference with a person based on her mistaken *values* in the supposed name of freedom itself. But need a defender of positive liberty include an external value requirement of this sort in her account of freedom? The tradition is speckled with thinkers who insist it does (as in Fichte's later thought). But what are the arguments for this? The issue turns, I think, on whether freedom can have the value attributed to it if it does *not* contain these normative components.²³ I would urge that it can. Moreover there is good independent reason to reject an external value requirement as one of the conditions for freedom. I will argue for this indirectly by critically discussing a recent attempt to defend a conceptual connection between freedom and particular value commitments.

In "Freedom and Value" Paul Benson argues that, in addition to self-control, there is a "further, equally significant ability necessary for free action. This is an ability to criticize courses of action competently by relevant normative standards."²⁴ Benson rests this claim on two arguments, one explicit and one less so. The explicit argument proceeds by the use of an example of a little girl who willfully plays where she is not supposed to but is too young to appreciate the value of refraining from her actions (she cannot appreciate the reasons underlying her parents' proscription of such actions). Benson claims that "in the sense in which

23. Gray, p. 337.

24. Paul Benson, "Freedom and Value," *Journal of Philosophy* 84 (1987): 465–86, p. 469.

we feel that an unqualifiedly free action is fully the agent's own, the child's action is not fully her own."²⁵ Since this is so the action is not free. Now it must be emphasized that Benson is not simply insisting that the person (child) have *some* notion of value that she can apply to her action. For certainly this child has that: she wants to play and puts little weight on the mess she is making and the inconvenience this causes. What Benson is claiming is that the child does not have an understanding of the *correct* norms applying to her action. By extension, then, the claim is that without proper attention to external (correct) norms of behavior, agents lack freedom.

This argument alone is rather weak, for it rests on highly disputable intuitions. I may well question the freedom of a child of the sort described for other reasons (e.g., that the desires that move her were not formed with the self-reflection necessary for freedom, or that the child lacked the ability to fully understand her actions and their consequences). But I would insist that it is not the failure to appreciate the *correct* norms that explains this lack of freedom. Certainly if we imagine the example as including an *adult* who willfully fails to appreciate the "correct" norm applying to a situation, we would not by that token cease to attribute freedom to her actions. Say the person is a confirmed egoist who refuses to abide *any* external values that refer to the interests of others. If such a person is not manipulated or deluded and has more or less consistent beliefs and desires, nothing about her grounds an intuitive judgment that she is lacking freedom.²⁶

Moreover, it seems downright counterintuitive to use the metaphor of "owning" one's actions as the motivating idea of this argument and to conclude that the only free actions available to the agent (the little girl, the egoist) are one's that are wholly external to her real motivations. The values and norms she *actually* holds do not give rise to free action on this account, where ones she precisely does not hold ("own") are the only ones that might. Thinking hard about ownership here as a central component of what it means to act freely supports intuitions exactly opposite to those Benson relies on.

This leads us to a less explicit argument that might be appealed to, which is central to many traditional defenses of positive liberty. This is that for freedom of agents to have the *value* it supposedly has—both for agents themselves and the rest of us looking on—'freedom' must mean, in part, conformity to generally accepted norms and values. The suggestion is that freedom has the value it has *only* insofar as it contains a requirement for conformity to the correct moral norms.

25. *Ibid.*, p. 478.

26. Consider the character Meursault in Camus's *The Stranger* who remorselessly commits murder. Such a person would be called unfree by theorists who demand an external value requirement for freedom, since he rejects all moral values in the performance of his acts. Others, though, would see his actions as the height of freedom (hence its frightening aspect).

But is this so? Certainly I value my ability to govern myself even if I believe that the values I am motivated by will turn out to be mistaken. While it may be an *additional* value to me that my free actions conform to what turn out to be the correct standards of value, it is hardly essential to the value of free action itself that my actions so conform. Imagine, for example, two agents who both perform similar morally wrong actions. One does so freely and deliberately and the second acts mindlessly, obediently carrying out the manipulative commands of her hypnotist master. Certainly the first agent is enjoying something of immense value that the second person lacks: the capacity for self-generation and self-government. And this is so despite the equally evil outcomes of both actions. To say, then, that freedom is a value only in relation to correct moral norms is to ignore the obvious noninstrumental value of self-mastery itself.

The mistake made here by critics of positive freedom, one to which defenders of certain versions of that concept have to some extent contributed, is that unless it is the *content* of the desires that are freely acted upon that is the locus of value of free action, then freedom has no value. What is ignored here is the special and intrinsic value that is contained in self-government itself. Just imagine being without it. What comes to mind as sorely lacking is *not* just the particular acts that could be carried out if it were returned, but the capacity for character formation and self-identity, which are themselves intrinsically valuable.

But, more strongly, if conformity to the "correct" values is a conceptually necessary condition for freedom then to the extent that this "correctness" is open to dispute or is indeterminate, then freedom of individuals is also indeterminate. Now this is not to invoke wholesale moral skepticism, but merely to point out the general variability and uncertainty that attaches to even the most foundational moral values. For example, which does freedom require that one give ultimate moral weight to: the maximization of utility or the rational autonomous will?

The point can be put another way. There remains a lack of stable consensus on a variety of deep moral questions. However, we don't take that as leaving in doubt the claim that a person in a particular case is acting freely. We don't postpone the question of free action until the moral controversy has subsided. Indeed the fact that the two questions are not even distantly related in our minds is evidence that the meaning of freedom is fixed independently of the determination of correct values. This, in turn, indicates that value commitments are not necessary conditions for freedom.²⁷

27. Benson is aware of this difficulty, but his reply is simply to claim that "there may be no such thing as the freedom of an action *simpliciter*. There may only be freedom in relation to specifiable domains of value" (p. 486). But this kind of relativism merely underscores, rather than tries to solve, the problem of the indeterminacy of the concept of freedom if one includes (external) moral requirements.

So for these reasons I conclude that there are adequate grounds for the rejection of the external value requirement for freedom of the sort that motivates the tyranny objection. So even if the internalist rationality conditions are accepted (plus the *epistemic* external requirement) as necessary conditions for positive freedom, the tyranny argument loses its force. The conception of positive freedom that emerges, then is a subjectivist, internalist notion, and one which adequately captures the core idea of self-government (and its value) without including untoward (external) value conditions.

What one is left with if one rejects the external value requirement (but accepts, in some defensible form, the other less stringent conditions) is what should be called a “content neutral” conception of positive freedom.²⁸ On this view, if the desires and values that a person develops are generated in accordance with the *procedural* conditions of autonomous preference formation that are constitutive of freedom, then no matter what the “content” of those desires, the actions which they stimulate will be (positively) free. There are good theoretical reasons for a content neutral conception. For any desire, no matter how evil, self-sacrificing, or slavish it might be, we can imagine cases where, given the conditions faced, an agent would have *good reason* to have such a desire. That is, there may be many cases where I freely pursue a strategy of action that involves constraining my choices and manipulating my values. But if this is part of an autonomous pursuit of a goal, it is implausible to claim that the resulting actions or values do not reflect my autonomy. So since we can imagine *any* such preference as being autonomously formed, given a fantastic enough situation, then it cannot be the *content* of the preference that determines its autonomy. It is always the *origin* of desires that matters in judgments about autonomy. This of course implies that some extremely constrained individuals will count as positively free because they (autonomously) choose to be under those constraints. But this is no more counterintuitive than the idea that I act freely when I have to shut off certain options to myself in order to achieve certain goals.

So while I take seriously the liberal objections to the traditional idea of positive liberty—a notion whose roots are in the political theory of the Romantic age—I do not regard this as the only understanding of the ideal of self-government. Rather than throw out the baby of self-government with the Jacobean bathwater, I am insisting on the development of a conception of *individual* positive freedom that avoids these worries of tyranny. What these arguments point to is that theories of justice should include central regard for such self-government in the concept of freedom which they defend. What is left of liberalism after such an inclusion, though, must remain for further discussion.

28. See Dworkin, *The Theory and Practice of Autonomy*, chaps. 1 and 2, who also defends a “content neutral” conception of autonomy.