

THE TROUBLE WITH FORMAL VIEWS OF AUTONOMY

Jonathan Knutzen

THERE IS a deep theoretical rift between formal and normative-capacity accounts of personal autonomy. According to formal accounts, personal autonomy consists in conditions that can be specified in purely structural or procedural terms.¹ According to normative-capacity accounts, personal autonomy consists, at least in part, in the possession of capacities for recognizing and responding to the norms that apply to one's choices and attitudes.² The first type of view denies that there are any substantive constraints on autonomously formed preferences and attitudes; the second affirms such constraints. In deciding on an account of personal autonomy, the choice between formal and normative-capacity accounts represents an important fork in the road: it is one of

- 1 Christman, "Autonomy and Personal History," "Liberalism and Individual Positive Freedom," "Procedural Autonomy and Liberal Legitimacy," and *The Politics of Persons*; G. Dworkin, *The Theory and Practice of Autonomy*; Ekstrom, "Autonomy and Personal Integration"; Frankfurt, "Freedom of the Will and the Concept of a Person" and "Autonomy, Necessity, and Love"; Friedman, *Autonomy, Gender, Politics*; Killmister, "Autonomy and False Beliefs"; Meyers, *Being Yourself* and "Decentralizing Autonomy"; Westlund, "Rethinking Relational Autonomy."
- 2 Benson, "Freedom and Value" and "Feminist Second Thoughts about Free Agency"; Kauppinen, "The Social Dimension of Autonomy"; McDowell, "Autonomy and Its Burdens"; Sayre-McCord and Smith, "Desires ... and Beliefs ... of One's Own"; Sher, *Beyond Neutrality*; Stoljar, "Autonomy and the Feminist Intuition." For a similar view about moral responsibility, see Wolf, *Freedom within Reason*. A brief note on terminology: in the recent literature on personal autonomy, these views are frequently called *normative competence accounts*. I prefer *normative-capacity accounts* instead because of the broader connotations of the label. Normative competence suggests psychological infrastructure that is relatively cross-situationally stable and purely internal to the agent. Normative capacity, by contrast, suggests something potentially broader, referring either to normative competence or to the ability to deploy that competence on this or that occasion. Since thinking of the normative abilities involved in personal autonomy in a wide, situation-inclusive sense is plausible, it is important to leave room for this interpretive option. Normative capacity therefore seems like a broader and potentially less misleading label for the family of views.

the deepest and most consequential choice points for our understanding of the nature of autonomy.

Formal accounts of personal autonomy represent the dominant type of view in the existing literature.³ It is not difficult to see why formal accounts have seemed attractive to many philosophers: they seem to avoid controversial assumptions about normativity and metaphysics and steer clear of undesirable political implications like perfectionism and paternalism. Notwithstanding their attractions, formal views have troubles of their own—troubles that are rarely noticed. As I will explain below, such views have difficulty making sense of the idea that autonomy entails a fairly robust form of responsibility, seem committed to an arbitrary asymmetry between the relevance of facts and values, and cannot properly vindicate the thought that autonomy is reason-giving in roughly the way we take it to be. By contrast, normative-capacity accounts have the resources to deliver a conception of autonomy that secures a robust form of responsibility, treats factual and evaluative information symmetrically, and shows why autonomy plays the normative role it does.

A full-scale comparison of the relative merits of formal and normative-capacity accounts would require considering all relevant theoretical costs and benefits of the competing views. Such comprehensive assessment is beyond the scope of this paper. What I offer instead is a *prima facie* case for reconsidering mainstream views of personal autonomy, highlighting some doubts about formal accounts and briefly indicating why the normative-capacity alternative need not be as troubling as it is sometimes taken to be. While I will not try to settle which type of account is ultimately superior, I do hope to contribute to clarifying what is at

3 Cf. Mackenzie and Stoljar, “Introduction,” 13; Westlund, “Rethinking Relational Autonomy,” 26. Note that formal views often go by other labels in the literature. Two common labels are “procedural” and “structural.” Perhaps the most common label for the type of view I have in mind is “procedural.” But there is precedent in the literature for calling the kind of view I have in mind “formal” (cf. G. Dworkin, *The Theory and Practice of Autonomy*, 12; Westlund, “Rethinking Relational Autonomy”), and I think doing so has certain advantages. The label “structural” is frequently associated with ahistorical views, like Harry Frankfurt’s, on which only the structure of an agent’s attitudes matter, whereas the label “procedural” is frequently associated with historical views, like Gerald Dworkin’s and John Christman’s, on which it matters how an agent’s attitudes come about and on which autonomy is centrally characterized in terms of actual or counterfactual psychological *procedures*. Some views, like Diana Meyers’s skill-based account, fit neither model but are nevertheless formal in the sense in which I intend. Since the labels “structural” and “procedural” have these more specific connotations that not all views in the family share, and since “formal” is a natural contrast to “substantive,” the term “formal” seems fitting for the broader family of views. However, I want to be clear that nothing I say below rides on this terminological choice. I will define formal views more carefully below. If one prefers the label “procedural” or “structural” for such views, then that is fine; the objections will be the same.

stake in the choice between these two alternatives. Any comprehensive assessment will need to begin by getting clear about the liabilities and advantages of each type of view. There are, as I will argue below, important and underappreciated systematic pressures favoring normative capacity over formal accounts of personal autonomy. If that is right, then appreciating these pressures will be an important step in assessing the merits of competing conceptions of personal autonomy.

The paper is in four sections. The first begins with a brief characterization of formal views. The following three sections articulate several *prima facie* objections to these views: that they cannot furnish an adequate account of responsibility, that they introduce *ad hoc* asymmetries between the importance of facts and values for autonomous agency, and that they are ill-equipped to vindicate the normative role played by the idea of autonomy. At the end of the paper, I briefly address why normative-capacity accounts need not have the troubling political implications some philosophers think they have.

1. WHAT IS A FORMAL VIEW OF PERSONAL AUTONOMY?

Formal views of personal autonomy come in many different shapes and sizes. The most popular variants offer some twist on the idea that to be autonomous one must be in some way identified with one's preferences and attitudes, for example, through taking reflective ownership of them or satisfying the condition that one would not disavow them if one became aware of their source.⁴ Some of these views also incorporate external conditions that must be met, e.g., that the formation of preferences occurs in the absence of coercion, manipulation, domination, and so on. Gerald Dworkin's classic statement of a formal view combines these two elements.⁵ On his view, personal autonomy consists in "a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values."⁶ Choices must issue from this capacity, but as Dworkin makes clear, they must also do so in conditions

4 Christman, "Autonomy and Personal History"; G. Dworkin, *The Theory and Practice of Autonomy*; Frankfurt, "Freedom of the Will and the Concept of a Person"; Friedman, *Autonomy, Gender, Politics*. Frankfurt's classic essay ("Freedom of the Will and the Concept of a Person") centers on the concept of freedom rather than autonomy. However, in a later essay ("Autonomy, Necessity, and Love"), Frankfurt extends his structural view of freedom to the concept of autonomy.

5 G. Dworkin, *The Theory and Practice of Autonomy*.

6 G. Dworkin, *The Theory and Practice of Autonomy*, 20.

of “procedural independence,” that is, in the absence of autonomy-undermining conditions like coercion and manipulation.⁷

The defining feature of formal views is that they exclude substantive elements by design. They do so at two levels.⁸ First, they place no *direct* constraints on the contents of choice: *any* choice can in principle be autonomously made. Second, they exclude *indirect* constraints on choice in the form of substantively defined attitudes or capacities featuring in the background. This means, for example, that on formal views autonomy cannot require that substantively defined attitudes, like self-esteem or self-respect, feature in the background of choice, as some philosophers have proposed.⁹ It also means that they cannot make substantively defined capacities, like the ability to appreciate and respond to genuine values and reasons, a condition of autonomous choice, as normative-capacity accounts maintain.¹⁰

To be sure, on many formal views, autonomy does require some kind of rational capacity. For example, it may require at least thin, procedurally defined, rational capacities like the ability to be sensitive to coherence constraints on beliefs and desires.¹¹ According to another popular suggestion, autonomy requires agents to be able to treat considerations *as* reasons. This thought can then be cashed out in functional terms. On a psychological version of the sugges-

7 G. Dworkin, *The Theory and Practice of Autonomy*, 16, 18, 20.

8 Cf. Benson, “Feminist Intuitions and the Normative Substance of Autonomy.”

9 Benson, “Feminist Intuitions and the Normative Substance of Autonomy.”

10 What I am here calling *direct* versus *indirect* substantive constraints is meant to be the same as Benson’s contrast between *strongly* and *weakly* substantive accounts of autonomy. The *direct*/*indirect* labeling is more intuitive. Suppose your view is that autonomy requires only choice in accordance with some narrowly defined substantive content: perhaps Adam and Eve are autonomous only so long as they do not eat the fruit on that one tree. And suppose my view is that autonomy requires capacities for reasons-responsiveness: to be autonomous, Adam and Eve must enjoy capacities for appreciating the reasons that bear on their choices. Then both of our views are substantive, but in different ways. Is one stronger or weaker than the other? That depends. One might very well interpret your view as imposing stronger conditions on autonomy than mine, or mine as imposing stronger conditions than yours. There is no obvious sense in which your view is substantive in a stronger sense than mine. The *direct* versus *indirect* distinction is clearer. What defines Adam and Eve’s autonomy on your view is that they make (or not make) certain choices. The relevant normative constraints thus bear a direct relation to the content of agents’ choices. What defines Adam and Eve’s autonomy on my view (and equally on views like Benson’s on which the agent must have certain self-regarding attitudes) does not depend on what they choose. Anything can in principle be autonomously chosen so long as certain background conditions upstream from choice are met. On this view, the relevant normative constraints bear an indirect relation to the content of agents’ choices.

11 Christman, “Autonomy and Personal History,” 14.

tion, considerations are treated as reasons when they play a specified role in the agent's psychic economy.¹² On a social version, considerations are treated as reasons when agents are prepared to answer for themselves in the interpersonal exchange of reasons.¹³ Crucially, however, on formal views there is no requirement that agents be, or have the capacity to be, attuned to *genuine* reasons. Indeed, there *cannot* be such a requirement consistent with the strictures of formalism. Since they are committed to doing without substantive commitments in specifying the criteria of autonomous agency, formal views are unhitched from objective values and reasons by design. Fidelity to the core commitments of formal accounts therefore requires that whatever rational facility is criterial of autonomous agency, it cannot be reasons-responsiveness in any sense that requires being hooked up to genuine and substantive normative features independent of the agent.

A caveat about this definition is in order. Formal views, I have suggested, rule out direct *and* indirect substantive constraints on choice. In some of the literature on personal autonomy, however, the focal contrast concerns only the first level: it is between views that are *directly* substantive and those that are not. For example, Dworkin contrasts his own view with a substantive account of autonomy, which he understands as placing *direct* constraints on what can be autonomously chosen.¹⁴ Similarly, Marilyn Friedman characterizes her view as "neutral with regard to the content of what a person must choose in order to be autonomous," and contrasts this with a substantive view according to which "someone is not autonomous unless she chooses in accord with certain values."¹⁵ The contrast invoked by Dworkin and Friedman does not perfectly align with the more demanding, two-level definition I have given.

Defining formal views in the more ambitious way is nevertheless appropriate for several reasons. First, it represents a trend internal to theorizing about autonomy by formal theorists themselves.¹⁶ This trend is in the spirit of non-substantive accounts of autonomy like those developed by Dworkin and Friedman, making explicit what these authors left implicit, or at any rate articulating a more thorough version of formalism. Second, the more ambitious two-level definition of formal views I have given follows more recent efforts at taxonomizing. The binary contrast invoked by Dworkin and Friedman does not adequately cap-

12 Bratman, "Intention, Practical Rationality, and Self-Governance."

13 Westlund, "Rethinking Relational Autonomy."

14 G. Dworkin, *The Theory and Practice of Autonomy*, 12.

15 Friedman, *Autonomy, Gender, Politics*, 19.

16 Christman, "Liberalism and Individual Positive Freedom"; Westlund, "Rethinking Relational Autonomy."

ture the space of interesting possibilities. For example, Paul Benson and Susan Wolf give accounts of autonomy on which it partly consists in the possession of reasons-responsive capacities.¹⁷ Benson has subsequently abandoned his earlier view and now defends the idea that autonomy requires certain substantively defined attitudes, like self-respect or a sense of self-worth.¹⁸ Both types of view have a claim to being substantive in an interesting sense, even if they do not require that an agent make specific kinds of choices. For good reason, therefore, both types of view are now routinely classified as substantive.¹⁹ This means formal views are best seen as those that exclude substance at the second and not merely at the first level, i.e., at the level of capacities and attitudes as well as the content of choice. Third, by focusing on a more thorough formalism, the two-level definition presents us with a sharpened and more interesting contrast. Few accounts in the literature are directly substantive in the way envisioned by Dworkin and Friedman—and for good reason. So far as I can see, such accounts do not seem highly compelling. By contrast, indirect substantive accounts of the kind proposed by Benson and Wolf have a good deal going for them. The binary contrast invoked by Dworkin and Friedman risks obscuring the most interesting and relevant alternatives.

Before moving on, it is worth clarifying that the critical upshot of some of the arguments presented below has relevance for views that are not strictly formal in the sense just defined. My criticism targets views that exclude normative capacity (or reasons-responsiveness) as a condition on autonomous choice. Consequently, insofar as the problems identified below are genuine, they will affect all views that exclude (or do not include) such capacities. Hence, views like Benson's, which require attitudes like self-trust or self-respect but do not require substantive normative capacities, are vulnerable to many of the criticisms identified below.²⁰ I nevertheless focus on formal views because these are popular and represent the starkest, most thorough alternative to thinking of personal autonomy in terms of the possession of normative capacities. They therefore

17 Benson, "Freedom and Value" and "Feminist Second Thoughts about Free Agency"; and Wolf, *Freedom within Reason*. Strictly speaking, these views are couched in the language of freedom rather than autonomy. But they nevertheless offer a model for a certain way of thinking about autonomy and are cited as such with some frequency in the literature on autonomy.

18 Benson, "Free Agency and Self-Worth" and "Feminist Intuitions and the Normative Substance of Autonomy."

19 Cf. Benson, "Feminist Intuitions and the Normative Substance of Autonomy"; Christman, "Autonomy in Moral and Political Philosophy"; Mackenzie and Stoljar, "Introduction"; Stoljar, "Feminist Perspectives on Autonomy."

20 Benson, "Feminist Intuitions and the Normative Substance of Autonomy."

constitute the most natural paradigm with which normative-capacity accounts can be contrasted.

With these clarifications in place, let us turn to some problems with formal views of autonomy.

2. THE RESPONSIBILITY CHALLENGE

A central problem with formal views of autonomy is that they cannot deliver an adequate conception of responsibility. In this section, I argue that there is a strong conceptual link between autonomy and responsibility and that reflecting on how best to interpret the notion of responsibility speaks against pure formal views and in favor of normative-capacity accounts.

Begin with the link to responsibility. The association between autonomy and responsibility is widespread in the literature.²¹ This is no accident; it reflects important conceptual connections between the two ideas.²² Whether it is made explicit or not, a basic assumption in much theorizing about autonomy is that autonomy is responsibility-entailing in roughly the following ways:

1. An autonomous agent is a responsible agent.
2. An autonomous agent is responsible for her choices and actions insofar as they issue from relevant autonomy-supporting capacities and circumstances.
3. All else equal, the greater one's autonomy in respect of choices and actions, the more one is responsible for those choices and actions.

Why think autonomy is responsibility-entailing in the sense expressed by these three claims? Let me highlight several pieces of evidence in support of this conclusion.

Performance respect. The exercise of autonomy capacities typically merits

21 For a small sampling: Arneson, "Mill versus Paternalism," 475; Buss, "Autonomous Action," 648; G. Dworkin, *The Theory and Practice of Autonomy*, 20; R. Dworkin, *Life's Dominion*, 224; Friedman, *Autonomy, Gender, Politics*, 21–22; Westlund, "Rethinking Relational Autonomy," 30–36.

22 Philosophers do frequently distinguish between autonomy and *moral* responsibility (e.g., Berofsky, *Liberation from Self*; Christman and Anderson, "Introduction"; Killmister, "Autonomy and False Beliefs"; McKenna, "The Relationship Between Autonomous and Morally Responsible Agency"). This shows that one can conceptually distinguish personal autonomy from moral responsibility. It does not show that there are no deep and interesting conceptual pressures linking personal autonomy and moral responsibility, much less that there are no deep and interesting pressures linking personal autonomy and a broader (less domain-centered) notion of responsibility.

a kind of appraisal respect related to the quality of an agent's performance.²³ When an autonomous agent enjoys circumstances conducive to the exercise of her autonomy capacities, she merits our esteem or disesteem on the basis of how she exercises her autonomy. I use the language of esteem and disesteem here because it has fewer moral connotations than praise and blame and leaves open the precise connection to moral praise and blame. (On some views, what I am here calling esteem and disesteem will turn out to be a kind of moral praise and blame.) What I have in mind is a credit-implying reactive attitude that tracks the exercise of capacities. While there are forms of appraisal that do not assume any notion of responsibility (e.g., appraising someone's physical attractiveness), other forms do, and it is quite plausible to think that exercises of autonomy are of this kind. At least in this context, esteem and disesteem are crediting responses, and they suggest that we see the agent as in some meaningful sense responsible for her choices and behavior. When we respond to a person with esteem or disesteem on the basis of how she exercises her autonomy, we plausibly see her as meriting such responses *via* her exercise of responsible agency.

The ethics of paternalism. It is plausible to think that at least part of what makes paternalism presumptively wrong is that it in some way violates autonomy.²⁴ In a suggestive metaphor due to Joel Feinberg, autonomous agents enjoy a kind of self-sovereignty.²⁵ Somewhat like the inappropriate meddling by one nation in the internal affairs of another, paternalistic interventions are thought of as illegitimate incursions into a person's proper sphere of choice. While one might interpret the sovereignty metaphor as suggesting that (hard) paternalism directed at competent adults can never be legitimate—that sovereignty sets an absolute side constraint—a weaker claim seems at least as plausible: competent adults are entitled to strong presumptions against paternalistic interference, making it difficult to justify warranted interferences for their own good. (Note that in the international arena, sovereignty is not plausibly absolute either.) The idea that competent agents are entitled to a sphere of choice is the idea of autonomy as a *right*. As I said above, how people exercise their autonomy capacities is associated with performance respect; by contrast, their right to make choices as they see fit is associated with recognition respect. To treat persons as little sovereigns

23 See Darwall ("Two Kinds of Respect") for the distinction between appraisal respect and recognition respect. Note that what I am here calling performance respect is only *part* of what Darwall calls appraisal respect. According to Darwall, appraisal respect includes assessments both of how agents perform in various roles/practices *and* of their characters.

24 Christman, "Autonomy in Moral and Political Philosophy"; Darwall, "The Value of Autonomy and Autonomy of the Will"; Feinberg, *The Moral Limits of the Criminal Law*; Groll, "Paternalism, Respect, and the Will."

25 Feinberg, *The Moral Limits of the Criminal Law*, 47.

is to treat them with due regard for their status as the kind of agents that merit protections against paternalistic interference.

This story seems to me hard to get off the ground without a background assumption about responsibility. Ideas about responsibility are plausibly implicated both in the scope and ground of the presumptive claim against paternalism. First, it is responsible agents who merit special protection against paternalistic interference. This is presumably why we think it is presumptively wrong to paternalize adults but not children. The difference is that adults are responsible agents whereas children are not. Second, facts about responsibility affect the case for and against paternalism, so that (all else equal) paternalism becomes harder to justify as responsibility increases and easier to justify as responsibility decreases. This is presumably part of the reason why it is much easier to justify soft paternalism. Think of Mill's classic example of a man about to walk over a bridge he does not know is unsafe. Paternalism is easier to justify in such a case than it is to justify in the case where the man, knowing the bridge is unsafe, intends to walk on it. This is so *whatever* one thinks about the all-things-considered justification of paternalistic intervention in the two cases. The case for paternalistic intervention is stronger in the first case than in the second, and a natural explanation for this is that facts about responsibility are salient: given his ignorance, the first man is less responsible for his choice (and the outcome of that choice) than the second.

If part of what makes paternalism presumptively wrong is that it violates autonomy, we have here a powerful reason to think autonomy is responsibility-entailing. To be sure, one could coherently accept that paternalism violates autonomy and that the case against paternalism is sensitive to facts about responsibility while denying any connection between the two. One might, for example, think paternalism involves the double wrong of violating autonomy *and* being inappropriately sensitive to facts about responsible agency. But this needs motivating. Antecedently, the simpler explanation connects the two: paternalism violates autonomy and autonomy entails responsible agency; it is the fact that agents are capable of being responsible for their own lives and choices that (in part) makes them autonomous; it is this very same fact that grounds a strong claim to being left free to pursue their lives and choices as they see fit. This picture is elegant in its simplicity, and it forges a straightforward connection between autonomy and responsibility.

Two brief caveats about the picture are in order. First, autonomy can imply responsibility without entailing that responsibility is sufficient for autonomy. There might well be additional elements to autonomy and, therefore, additional wrong-making features to paternalism. It is common, for example, to distinguish autonomy as a right from autonomy as an agency ideal. Plausibly, autonomy as

an agency ideal is more demanding than mere responsible agency. But it nevertheless entails responsible agency. For the ideal to be in the offing, one has to be a responsible agent, capable of being responsible for one's life choices in some suitably rich and meaningful sense. Second, a well-justified regime of anti-paternalist norms presumably has multiple sources of justification.²⁶ My point is not that considerations of responsibility are the exclusive source of justification for anti-paternalist norms, but that they are one important plank in the ethics of paternalism. Moreover, it is only fair to acknowledge that justifications of anti-paternalist norms are conceivable that make no appeal whatsoever to responsible agency (e.g., that agents typically know best what is in their interest). To fully defend the claim that strong anti-paternalist norms are best justified by a background assumption of responsible agency would require showing that alternative justifications, which do without the assumption of responsible agency, are not sufficient. That is more than I can do here. I will therefore content myself with making the bet that these alternative explanations fail. They may contribute to partial justifications for anti-paternalist norms, but it is doubtful that they can deliver complete and adequate justifications.²⁷

When someone decides to smoke or climb dangerous mountains, that choice plausibly merits respect *as* the choice of a responsible agent. Hence, as noted above, when someone is adequately informed about the risks, there is also stronger reason to desist from interfering. This is a backward-looking responsibility rationale: all else equal, there is more reason to allow persons to reap the consequences of their actions when they are undertaken responsibly than when

26 Thanks to Richard Arneson for helping me to see this more clearly.

27 For example, epistemic arguments are fragile. While it may be true that people tend to be better judges of what is in their interest because they have access to a richer base of relevant information, that is not always so. Moreover, many of us want to say that paternalism can wrong someone even if we grant that the paternalizing agent knows better. Another candidate explanation for the presumptive wrongness of paternalism, in many ways elegant in its simplicity, is that people just *do not like* being forced to do things against their will. (Thanks to an anonymous reviewer for *JESP* for raising this possibility.) As with the epistemic argument, this type of explanation may contribute to a general anti-paternalist rationale, but I doubt it can be the whole story. Strong-willed agents do not merit greater anti-paternalist protections than docile agents; very small and non-responsible stubborn children do not merit stronger anti-paternalist protections than equally non-responsible but agreeable children. Moreover, the reasons given by an agent's desires seem not to ground the full-orbed recognition respect that seems appropriate to an adult chooser. To be sure, honoring people's wishes so far as possible is plausibly a way of respecting them. However, the recognition respect appropriate to an adult chooser seems to require more: the recognition not only that here is an individual with desires and a perspective on the world (a child is an agent of this sort), but that here is a *responsible* agent—an agent whom it could make sense to let live with the potentially momentous consequences of her choices.

they are not. One might prefer a more forward-looking responsibility rationale instead.²⁸ Perhaps a regime of anti-paternalist norms can be partly justified by its proleptic or educative effects, tending to cultivate capacities for responsibility of roughly the kind it appears to assume. Either way, without the idea that persons are, or can become, responsible choosers, it is very difficult, I think, to support quite robust and general anti-paternalist presumptions of the sort most people in liberal societies subscribe to. The point here is not that people always live up to this picture of responsible agency or that facts about responsibility entirely settle issues about the ethics of paternalism. Rather, it is that our commitments to anti-paternalist norms plausibly depend on a deep background assumption of responsible agency.

Options. The third line of evidence comes from the persistent attractiveness of the idea that options matter for personal autonomy.²⁹ Raz gives memorable examples.³⁰ The man who falls into a pit and can only decide when to nap or which direction to move his head is not very autonomous. Nor is the woman who is trapped on an island with a hungry beast and who spends her every waking moment trying to avoid being eaten by it. Something similar goes for the slave, who lacks options and cannot choose his own course through life, as well as for the many more prosaic forms of impoverishment that may not involve domination but nevertheless involve restricted options.³¹ For example, it seems natural to describe refugees trapped in refugee camps as suffering diminishment of autonomy.³²

Lack of options is constraining; it leaves agents less free to choose their course. It thereby also tends to diminish responsibility. Those who lack adequate options will tend to be less responsible for their choices and for the consequent shape of their lives.³³ The man in the pit is responsible for a few things—for whether he naps now or later, for which way he turns his head. But he is not responsible for much else about his life. His constrained circumstances change how it is appropriate to appraise the man. Before he fell into the pit, it might have

28 Cf. Vargas, *Building Better Beings*.

29 Hurka, "Why Value Autonomy?"; Kauppinen, "The Social Dimension of Autonomy," 284–86; Mackenzie, "Three Dimensions of Autonomy," 28; Mackenzie and Stoljar, "Introduction," 22, 26; Oshana, "Personal Autonomy and Society," 94, and *Personal Autonomy in Society*; Raz, *The Morality of Freedom*, 373–77; Terlazzo, "Conceptualizing Adaptive Preferences Respectfully."

30 Raz, *The Morality of Freedom*, 373–74.

31 Oshana, "Personal Autonomy and Society"; Nussbaum, *Creating Capabilities*; Sen, *Development as Freedom*.

32 Betts and Collier, *Refugee*, ch. 6.

33 Cf. Hurka, "Why Value Autonomy?"

been appropriate to feel some disesteem for him because, while enjoying significant talent and opportunities, he spent most of his days playing video games. Now that he is trapped in the pit, however, it would be absurd to feel disesteem for him on the basis of his unambitious choices. Because he lacks opportunities to exercise his agency capacities in a meaningful way, such performance-based assessments would be out of place. The impact of limited opportunity on *moral* accountability is familiar from fair-opportunity accounts of moral responsibility.³⁴ Something similar seems plausible in the case of personal autonomy. In general, lack of options tends to spell diminishment of autonomy. If personal autonomy implies responsibility, we can make sense of this. Limited opportunity undermines or threatens autonomy because, all else equal, it makes persons less responsible for their choices and lives.

Self-authorship/self-creation. A final piece of evidence for the link between autonomy and responsibility is to be found in widespread appeals to tropes of self-authorship and self-creation throughout the autonomy literature.³⁵ These metaphors express something deep and important about what it means to be autonomous, yet they are hardly intelligible without some background idea that persons are responsible for their lives. Creators and authors, after all, must be more than merely *causally* responsible for the products they create or author. To be self-authors or self-creators in any meaningful sense, persons must enjoy the right kind of responsibility-conferring relationship to their choices and lives.

Together, these four lines of evidence suggest significant connections between personal autonomy and responsibility. In particular, they suggest that it is plausible to think of autonomy as responsibility-entailing in roughly the way suggested: that to be an autonomous agent, one must be a responsible agent; that an autonomous agent is responsible for her choices and actions when they issue from favorable circumstances; and that, all else equal, greater autonomy in respect of choices and actions implies greater responsibility for them. If the connections we have noticed are genuine, it is little wonder that the idea of responsibility crops up with some frequency in discussions of autonomy. There are systematic pressures supporting the idea that autonomy is responsibility-entailing.³⁶

34 Brink and Nelkin, "Fairness and the Architecture of Responsibility."

35 E.g., Benn, "Freedom, Autonomy and the Concept of a Person," 125, 127; R. Dworkin, *Life's Dominion*, 224; Enoch, "Hypothetical Consent and the Value(s) of Autonomy," 27; Griffin, *On Human Rights*, 150; Raz, *The Morality of Freedom*, 369–70, 390.

36 An anonymous reviewer from *JESP* asks a helpful clarifying question: Does the entailment run only in one direction or in two? And what exactly is the intended order of explanation here? Answer: what I am suggesting is that responsible agency is a necessary but not a sufficient condition of autonomous agency. If we imagine a Venn diagram, the circle labeled

Now for the trouble. Many formal theorists accept that autonomy comes with responsibility. Consider a representative quote from Gerald Dworkin: “By exercising [their capacities for autonomy], persons define their nature, give meaning and coherence to their lives, and take responsibility for the kind of person they are.”³⁷ The question is whether they have the resources to make sense of this commitment. More specifically, the question is whether they can deliver a notion of responsibility that is adequate to the task. Formal views spell out the conditions of autonomy in terms of properties like structural mesh between attitudes of higher and lower orders, agential coherence, actual or counterfactual reflective endorsement, forming temporally extended plans, treating considerations as reasons in the evaluation and adoption of plans, and so on. Such properties do seem well-suited to furnishing the basis for *some* ascriptions of responsibility. In particular, they seem to support judgments of attributability, according to which agents are related to their actions in such a way that their actions manifest their character and commitments.³⁸ In the case of moral conduct, responsibility-as-attributability typically means that actions express an agent’s quality of will. But the idea of attributability can be generalized to cases not limited to moral matters. An agent will be attributively responsible for her life choices (even purely self-regarding ones) if they reflect on her—on what kind of person she is, on her sense of self, on her character, priorities, commitments, and values. Many formal views of personal autonomy are preoccupied with authenticity conditions. These aim to tell us when some choice or attitude is the agent’s own in a special sense. Such accounts therefore seem to be well-equipped to capture

“autonomy” would be entirely within the circle labeled “responsibility”—the one is a subset of the other. On this picture, autonomy entails responsibility but not vice versa. Note that this is a *logical* claim. On a plausible interpretation, the supporting metaphysics that would make this logical claim true would involve a claim about constitution: autonomy is constituted, at least in part, by responsible agency. This makes responsibility (at least in one sense) *explanatorily* prior.

37 G. Dworkin, *The Theory and Practice of Autonomy*, 20.

38 Following Gary Watson (“Two Faces of Responsibility”), many discussions of moral responsibility distinguish two senses of responsibility: attributability and accountability. Roughly, one is responsible in the attributability sense if one’s actions reflect one’s quality of will, and one is responsible in the accountability sense if one’s actions meet whatever control conditions are required for being held morally accountable. A plausible specification of the control conditions involved in moral accountability is that they consist (at least in part) in the possession of normative capacities (Brink and Nelkin, “Fairness and the Architecture of Responsibility”; Fischer and Ravizza, *Responsibility and Control*; Nelkin, *Making Sense of Freedom and Responsibility*; Wolf, *Freedom within Reason*). By contrast, a plausible specification of attributability-relevant conditions requires only that an action reflect something like the agent’s genuine or authentic self—her character, perspective, or will.

the sense in which people can be attributability-responsible. When agents meet the requisite authenticity conditions, they stand in the relation of ownership to their choices and attitudes such that those choices and attitudes reveal where the agent stands, what she is about, and so on. Such views can therefore yield an important sense of responsibility: the kind that reveals something of the agent's inner life, putting her on display and opening her up to certain forms of appraisal.

The crucial question is whether this conception of responsibility is the right kind. Is it adequate for an account of personal autonomy? Two considerations suggest it is not.

First, as we have seen, the exercise of autonomy capacities typically merits performance respect. Since formal views of autonomy can plausibly make sense of attributability-responsibility, they plausibly have the resources to make sense of certain forms of appraisal respect: character-grading, aretaic judgment, assessment of motive, and so on. But performance respect requires something more specific. When an agent merits our respect for the exercise of her autonomy, her actions must meet a kind of credit condition such that the agent can earn our esteem or disesteem on the basis of how she exercises her autonomy. Some crediting responses are quite weak: they amount only to something like approval or disapproval. Attributability responsibility suffices for making this weak class of responses apposite. Other crediting responses, however, are stronger: they amount to something like performance criticism. It is not clear attributability responsibility suffices for this stronger class of responses.

Consider that the facts determining choice-worthiness are normative. Since the paradigm of personal autonomy is often taken to be self-regarding choice, consider for simplicity the domain of prudence. On all of the most widely held and plausible views of welfare, there are facts about what is good for agents that is independent of their momentary desire and whim. Choices in this domain can be better and worse, right and wrong, wise and unwise, and so on.³⁹ It is hard to see how the stronger class of crediting responses could be apposite in the absence of sensitivity to the very facts that determine choice-worthiness. To be a suitable target of performance criticism on the basis of how an agent exercises her autonomy capacities, she must enjoy the right kind of control. But it is hard to see how the agent could enjoy such control in the absence of normative capacities. Agents who satisfy formal autonomy criteria but lack normative capac-

39 This is true on hedonist, objective-list, and perfectionist views, but it is true on the most compelling versions of the desire-satisfaction view as well, which add counterfactual and idealizing conditions as a filtering mechanism on desires, the fulfillment of which count toward a person's welfare.

ities seem a bit like blindfolded dart throwers attempting to hit a target.⁴⁰ Why should an agent merit our disesteem if she is completely insensitive to the facts in virtue of which she ought to choose one way or the other or cannot suitably regulate her conduct in light of this sensitivity? Unhitch agents from the relevant normative facts, either because they are blind to them or incapable of acting on them, and it becomes very hard to see how they can be responsible for their choices in the way that is characteristic of the kind of performance respect we associate with the exercise of autonomy.

Second, as we also have seen, on a plausible interpretation of the ethics of paternalism, both the scope and grounding of anti-paternalist principles are sensitive to facts about responsible agency. Does being attributability-responsible suffice to ground robust anti-paternalist norms? It is hard to see how it could. The same facts that make it difficult to see how an agent who lacks normative capacities could have the kind of control needed to render performance criticism apposite also make it hard to see how it could ground a strong claim against paternalistic interference: it is precisely because children lack such capacities that they do not have a strong claim against intervention by parents and educators.

Formal views of autonomy do frequently posit reflective ownership capacities. Would such capacities suffice to merit anti-paternalist protections? It is hard to see how. Again, the domain of choice is governed by practical norms. Stipulate that the agent is insensitive to these norms and it becomes difficult to see why she merits strong protection against intervention by third parties. To be sure, since facts about responsibility are not the only relevant facts for determining the appropriateness of paternalistic intervention, there may still be all-things-considered reasons to protect her choice even if she is not sufficiently responsible. But, as I argued above, facts about responsible agency are a huge pillar in the anti-paternalist case. Once this is acknowledged, we need an interpretation of the relevant notion of responsibility. What we need is a kind of responsibility that is robust enough to ground strong anti-paternalist norms and (as a corollary) puts agents on the hook for the upshots of their own choices. Mere identification with, or ownership of, attitudes seems much too weak. What matters is not that a choice is authentically yours; what matters is that the choice is one for which you are robustly responsible, in such a way that you can be on the hook for its consequences, and that I have strong presumptive reasons to let you make your choice even if I could do better. To secure this result, something more is needed. Adding reflection does not do the trick. Perhaps reflective endorsement increases authenticity, such that you then own the choice in a special and

40 For the metaphor of blindness, see Kauppinen, "The Social Dimension of Autonomy," 281; Wolf, *Freedom within Reason*, 92.

deeper way. This may say something about you—about your character, perspective, and values. In that minimal sense, it constitutes a kind of responsibility. But reflective endorsement does not amount to enjoying a relation to your choices that would explain why other agents have presumptively decisive reasons to let you have your way, even when they know better. The fact that a choice is *authentically yours* just does not seem to have the right kind of normative relevance to ground such reasons. By contrast, if you have normative capacities for appreciating and responding to the values and reasons bearing on your choice, then that does seem to do the trick. If you can appreciate and respond to the normative features relevant to your choice, then that gives you a deeper kind of control over your choices and actions, putting you on the hook for their upshots, and giving me reasons to desist from paternalistically interfering with your choice.

To return to the difference between children and adults: What is the salient difference between children and adults, such that paternalism of the former is generally more acceptable than paternalism of the latter? It is plausibly a difference in their status as responsible agents. But what kind of responsibility is relevant here? I have argued that strong anti-paternalist protections would be better supported by a form of responsibility that puts agents on the hook for the upshots of their choices than a form of responsibility that merely reveals what kind of person they are, where they authentically stand, etc. If that is right, the relevant difference between children and adults seems to be that the former have more fragile normative capacities than the latter, not that they have a less well-developed sense of self. This has some intuitive plausibility. Think of it this way. You are about to meet a seven-year-old child who is a stranger to you. All you know is that the child is extremely precocious and wants to undertake a dangerous activity. You have the power to stop her. Which set of facts would ground a stronger claim against you not to interfere with her choice? The fact that she is reflectively mature and seems to have crystalized a perspective and stance on the world that is genuinely her own? Or the fact that she is reflectively mature in such a way that she seems sensitive to normatively relevant features of her choice? I suspect you will agree that normative capacities ground stronger claims against intervention than mere authentic ownership of choices. While children generally have more fragile normative capacities and a less developed sense of self, it is the first of these properties that seems more important in considering whether paternalistic treatment is warranted.

Together, these considerations put enormous pressure on formal views of autonomy. What we need, I have argued, is an interpretation of autonomy that can deliver a robust conception of responsibility. The worry is that formal views cannot deliver such a conception. They can give us a conception of responsibil-

ity that shows us where the agent stands and thereby reveals something good or bad about her. But they cannot give us a conception of responsibility that shows the agent to be in the kind of relationship to her attitudes and choices that seems to be required by our treating exercises of her autonomy as meriting positive or negative performance respect; nor can they give us a conception of responsibility robust enough to ground strong claims in favor of allowing the agent to live with her choices and against others that they not paternalistically interfere.

3. THE FACT/VALUE ASYMMETRY

Perhaps formal theorists will succeed in giving us a rich and convincing story about responsibility. Even so, a further problem looms. Responsibility requires adequate non-evaluative information. This idea is familiar from discussions of *moral* responsibility, where ignorance is typically taken to be exculpatory. Ignorance, of course, does not *always* excuse, as in the case of willful or negligent ignorance, but ignorance can and frequently does serve as an excusing condition in assessments of moral responsibility. The analogous thought is plausible in the case of personal autonomy as well: just as inadequate information can diminish moral responsibility, inadequate information can diminish personal autonomy. Someone who smokes in complete ignorance of the risk this poses to her health is plausibly less autonomous with respect to that choice than someone who is apprised of the facts and chooses to smoke anyway; a lover who marries her beloved ignorant of his true character is less autonomous with respect to that choice than someone who knows her lover in greater depth; and so on. All else equal, more choice-relevant information means more autonomy; less choice-relevant information means less autonomy.

It is possible to deny that non-evaluative information is relevant to autonomy. Michael McKenna commits himself to this bold thesis in an effort to discover some interesting differences between moral responsibility and personal autonomy.⁴¹ In his central example, Tal attempts to help his sick friend, Daphne. Pulling mislabeled medicine from the cabinet, Tal gives Daphne poison and thereby accidentally poisons her. According to McKenna, Tal acts autonomously (though he is not morally responsible) in poisoning his friend. This is because, explains McKenna, there is a sense in which Tal rules himself by acting in accordance with self-chosen principles. The principle on which Tal acts is: always attempt to help those who suffer innocently. And Tal's action conforms to this self-chosen principle because, in administering the drug, he does attempt to save his friend.

41 McKenna, "The Relationship Between Autonomous and Morally Responsible Agency."

The resulting picture is a fairly stark form of internalism on which autonomy is compatible with sweeping ignorance of relevant facts.

This is not compelling. On its own, the example seems to lend intuitive support to the thought that Tal's autonomy is undermined or threatened by his ignorance. So do similar examples discussed by Al Mele, like the example of Connie, who chooses an investment plan but is systematically deceived by the company offering the plans, and King George, who rules his kingdom contrary to his deepest commitments because his staff systematically distorts the information arriving at his desk.⁴² If anything, it seems intuition antecedently favors the verdict that these agents suffer some impairment of autonomy by being informationally cut off. McKenna acknowledges the intuitive pull of Mele's examples, but he insists that the intuitive pull tracks moral responsibility rather than personal autonomy. If we stipulate that autonomy is acting in light of self-chosen principles, McKenna suggests, Tal and Connie and King George can all be seen as autonomous. But this is not obvious. Even granting McKenna's stipulative definition of autonomy, no strong form of internalism follows, for it is plausible to suppose that acting in light of one's principles imposes success conditions on action that are not met in the examples.⁴³ McKenna avoids this problem by describing Tal's action (and by implication, Connie's and King George's) as an *attempt*. Tal's principle is: *attempt* to help your friends. This is something he succeeds in doing. But suppose his principle were: help your friends. This is not something he succeeds in doing. The intuitive force of the examples as instances of autonomy thus depends on an artefact of description. Tal and Connie and King George would, of course, not be identified with their actions under an informationally enriched perspective. Once we shift the act description to a more objective frame, it becomes much less compelling to think of them as autonomous. Consider another possible fix. One might describe all principles of action in evidence-relative terms. Tal's principle might be: act to fulfill your goals and values as indicated at the moment of action by your subjective evidence base. He might then, for example, grievously harm Daphne while counting as exemplary in his autonomy simply because he acts on his (misleading) evidence about what helps and harms her. But what would serve the values and principles of agents is not typically evidence-relative or subjective in this way: Tal cares about his friend, Connie cares about her future in retirement, King George cares about the flourishing of his kingdom, and so on. This outward focus imposes objective success conditions that require being suitably well-informed if one is to promote the relevant values and principles. Antecedently, as I said, intuition seems to fa-

42 Mele, *Autonomous Agents*, 179–81.

43 For similar points of criticism, see Killmister, "Autonomy and False Beliefs."

vor Mele's verdict about the cases, that autonomy is threatened by deprivation of decision-relevant information. One can try to deflate some of the intuitive force of these examples, as McKenna seeks to, but only by redescribing the principles and values from which agents act in terms of implausibly unambitious success conditions—as attempts or evidence-relative respondings. If one sticks with a realistic interpretation of what agents actually care about, then their being autonomous plausibly does depend on being adequately informed.

As we have seen, there are strong conceptual and theoretical pressures to preserve the association between personal autonomy and responsibility. McKenna arrives at his conclusion precisely in an effort to locate some interesting notion of personal autonomy that *comes apart* from responsibility. But he offers neither theoretical motivation nor robust conceptual anchor points for this strong internalist suggestion. If, as I argued above, autonomy is a form of personal freedom in virtue of which agents are responsible, then we have good reasons to reject the kind of extreme informational hermeticism on which an agent can be completely ignorant or deceived about factual information relevant to her choice.

Theoretical pressure is increased by noticing the connection between being informed and having control. Ignorance threatens an agent's control.⁴⁴ The examples of Tal, Connie, and King George exemplify this. By being significantly ignorant, these agents have impoverished control over their actions. And control seems fairly clearly relevant to autonomy. Think of a case involving complete absence of executive control. Perhaps in the inner sanctum of my mind I endorse normative principles and aim to conform my actions to them. It seems utterly implausible to think that I enjoy autonomy if I have no power whatsoever to conform my actions to my principles. But if lack of control threatens autonomy on the "active" side, why would it not do so on the "receptive" side as well? After all, both executive capacities and representational capacities can be thought of as aspects or dimensions of control. In the absence of reasons to posit an asymmetry in control conditions, it seems arbitrary and unmotivated to insist that one dimension of control matters while the other does not. In short, McKenna's proposal is an interesting suggestion about how to divide conceptual labor between moral responsibility and personal autonomy, but we have few independent reasons to accept it and some independent reasons to reject it.

Must we go to the other extreme and hold that only those who act in light of all relevant information are autonomous? This would entail the absurd conclusion that almost no one ever acts autonomously. We need not accept this extreme conclusion. An intermediary view is available, namely that *sufficient* information is necessary for autonomy. This is plausible if we distinguish scalar and

44 Cf. Mele, *Autonomous Agents*.

threshold assessments of autonomy. Many of the properties relevant to autonomy (like being informed) are a matter of degree. Deploying a scalar conception of autonomy, we can say that agents are more autonomous the more they satisfy the relevant scalar property. Switching to a binary, threshold conception, we can say that some threshold level of the property must be reached to qualify as autonomous. The two conceptions can be combined. On such a picture, autonomy “kicks in” above the threshold but one can be more or less autonomous (perhaps with no upper bound) above that point. When it comes to being informed, the combined conception seems plausible. Below some threshold of understanding, agents may not be autonomous with respect to a choice at all. Above that threshold, being more informed tends to enhance, and being less informed tends to diminish, autonomy. Citing Columbus’s ill-informed decision to sail west, Nomy Arpaly doubts “that anyone wishes to claim . . . that an ill-informed decision cannot be an instance of autonomous agency.”⁴⁵ But she also accepts that giving someone more information might make the person more autonomous. Once we distinguish scalar and threshold verdicts, both of these claims seem plausible. Being ill-informed can be autonomy-impairing while not rendering one entirely non-autonomous.⁴⁶

Now for the challenge. Suppose formal views accommodate the idea that non-evaluative information is relevant to autonomy. By parity of reasoning, it seems plausible to suppose that autonomy likewise requires *normative* information. It is hard to see why someone who is completely normatively “blind” would be any more autonomous than someone who is ignorant of non-evaluative information. Suppose someone smokes, knowing the risk this poses to her health but in total ignorance of what is good for her or the reasons this gives her to make one choice rather than another. Such a person seems just as blind, in the relevant sense, as someone who is ignorant of the non-evaluative facts. There

45 Arpaly, “Responsibility, Applied Ethics, and Complex Autonomy Theories,” 175.

46 So far as I can see, the *source* of misinformation is irrelevant to autonomy. Tal’s autonomy is not lessened more if his misinformation is the result of intentional manipulation than if it is the result of accidental labeling. Similarly, if Connie and King George are informationally impaired due to a fluke of circumstance, this is no less an impairment of their autonomy than if they are the victims of campaigns of disinformation. To be sure, we are likely to see social sources of autonomy deficits as having special moral significance. We can, for example, condemn (unjustified) manipulation and blame those who engage in it. But ignorance is ignorance and seems threatening to autonomy no matter what its source. If my autonomy is threatened by your intentional deception of me, there is little reason to think it would be less threatened if my ignorance is the result of impersonal forces. Of course, this is holding all else equal. Social sources of misinformation may be worse threats to autonomy insofar as one person is in the power of another. If your manipulation is a way of also dominating me, then the way in which my ignorance originates seems more significant.

is, then, a simple parity argument for treating factual and normative ignorance alike.⁴⁷ If autonomy implicates responsibility, and if both factual and normative ignorance can defeat or attenuate responsibility, we have (in the absence of further considerations) no more reason to credit autonomy in the absence of the one than in the absence of the other.

Extreme internalist conceptions of autonomy are implausible. They suggest that complete factual ignorance does not in any way threaten autonomy. Formal theorists therefore do well to accept that non-evaluative information can make a difference to autonomy.⁴⁸ But once this much is accepted, there is pressure to accept that evaluative information is relevant to autonomy as well. If one accepts that autonomy is responsibility-entailing, there is a principled rationale for taking this further step. Sensitivity to evaluative information is just as relevant as purely factual information in constituting an agent as responsible. The domain of choice is one in which norms apply: choices can be better or worse, right or wrong, prudent or imprudent, and so on. Truths about choice-worthiness are a function, not of descriptive facts *per se*, but of descriptive facts *plus* relevant evaluative or normative truths. Hence, truths about choice-worthiness are partly normative. But truths about choice-worthiness also furnish the basis for critical assessments of agents. Factual and normative deficits alike tend to be responsibility-diminishing: below some minimal threshold level, agents are not responsible for their choices at all; above that level, they are more or less responsible, depending on their sensitivity to the relevant features.

Formal views, however, must reject parity. Since they sever the connection between autonomy and substantively defined evaluative capacities, such views must also deny that normative information matters for autonomy.⁴⁹ This creates an explanatory burden. On the face of it, the exclusion of evaluative information seems *ad hoc*. This puts pressure on formal accounts to explain why evaluative and non-evaluative information should be treated in an asymmetric fashion. The answer cannot be: because that is what is predicted by formal theories. In the absence of some salient difference we have independent reasons for accepting parity. This speaks against formal theories precisely because they predict an asymmetry. It is therefore not satisfactory to point to this implication of formal

47 On the symmetry of facts and values in moral responsibility, see Rosen, "Culpability and Ignorance"; and Wolf, *Freedom within Reason*. For the parallel claim about personal autonomy, see Kauppinen, "The Social Dimension of Autonomy," 280; and Savulescu, "Rational Non-interventional Paternalism," 330.

48 Cf. Berofsky, *Liberation from Self*; Killmister, "Autonomy and False Beliefs"; Mele, *Autonomous Agents*.

49 Killmister does just this, arguing that false factual beliefs tend to impair autonomy but not false principles or values ("Autonomy and False Beliefs," 527).

theories in reply to the challenge. Perhaps formal theorists can ultimately give us some principled, non-question-begging story about why we should treat normative and factual information differently. In the meantime, we have a *prima facie* case for thinking autonomy requires sensitivity to evaluative information. This speaks against formal views of autonomy.

4. AUTONOMY'S NORMATIVE ROLE

The final worry about formal views is that they cannot make adequate sense of autonomy's normative role. We recognize autonomy's normative role in the kinds of reasons it supplies. Autonomy is reason-giving in roughly two ways. On the one hand, we think it good, all else equal, for people to live autonomous lives and make choices autonomously. It is therefore the sort of thing we have reason to aspire to ourselves and promote the realization of in others. On the other hand, we think autonomy marks out a sphere within which individuals are free to choose and that their autonomous choices carry authority or gravity in certain contexts of decision-making to which we must often give greater weight than the choices would merit on the basis of their direct consequential value or other forms of choice-worthiness.⁵⁰ When an agent or her choices meet the conditions of autonomy, we must take her decision with special seriousness. Even when it is trumped by other considerations, autonomy places the bar of interference higher than it otherwise would be: it ratchets up the demands for warranted intervention.

An adequate conception of autonomy should be able to make sense of this twofold normative role. In other words, an adequate conception of autonomy needs to vindicate the thought that autonomy is worthy of promotion and worthy of respect. But there are reasons to doubt that formal views provide ingredients sufficient to meet this demand. Let us consider each of these normative roles in turn.

What are the kinds of autonomy-relevant conditions we generally have reasons to promote? The most obvious is perhaps this: to ensure that people have sufficiently valuable options to choose from. Some autonomy theorists, like John Christman, deny that valuable options matter for autonomy.⁵¹ But often the

50 One might object to putting the point in terms of the weight of reasons. According to Groll, an autonomous will is to be taken as "structurally decisive" ("Paternalism, Respect, and the Will," 699–706). However, Groll suggests that paternalism is only "presumptively wrong" (710–11). So even on a Raz-style view like Groll's, where certain considerations are shielded from entering deliberation, the shield is not necessarily absolute.

51 Christman, "Procedural Autonomy and Liberal Legitimacy," 282. Christman subsequently argues that if valuable options matter to autonomy this should be understood in terms of a

motivation for this denial is heavily theory driven, for example, by the desire to avoid perfectionist implications. Pre-theoretically, it is quite natural to describe people with limited valuable options as suffering a diminishment of autonomy.⁵² Consider refugees living decades of their lives in a camp. These people typically have a dearth of valuable options and it is natural to think of them as suffering from an autonomy-relevant impairment as a consequence.⁵³

Valuable options are an external good. We plausibly also have reasons to promote an internal good to go along with it. Think of what parents want for their children. Parents do not just want their children to face a lush banquet of valuable options; they want their children to possess the capacities to appreciate and respond appropriately to those options. This pattern of concern seems appropriate more generally. It would seem a bit odd to care that persons enjoy valuable options but not to care that they enjoy capacities for appreciating and responding appropriately to those valuable options. Some formal theorists argue that autonomy requires valuable external options but that the internal capacities should nevertheless be understood in terms of purely procedural conditions.⁵⁴ This is an unstable position. Once one accepts that valuable options matter, why not also accept that internal competencies for tracking and pursuing those valuable options matter?

There are, of course, complicated questions about who may promote whose autonomy and how this may be done. Some liberal theorists, for example, insist that the state may play no role in promoting the autonomy of its citizens. Moreover, there are plenty of cases where we have reasons not to promote, and even to curtail, autonomy—for example, prospectively, when people's exercise of autonomy will likely bring about significant and unjustified harms to others, and retrospectively, for purposes of punishment. But the present point does not depend on denying such qualifications. What it depends on is only the broad generalization that people ordinarily have robust reasons to promote their own autonomy and often also the autonomy of others. The exceptions are important, but they should not obscure the fact that there are general standing reasons for anyone to promote anyone else's autonomy. A plausible interpretation of what people generally have reason to promote includes (i) valuable options and (ii) normative competence over those options. It is consistent with this to think that

subjective conception of value, i.e., the options need only be valuable from the perspective of the agent (*The Politics of Persons*, 170).

52 Cf. Nussbaum, *Creating Capabilities*; Oshana, "Personal Autonomy and Society" and *Personal Autonomy in Society*; Raz, *The Morality of Freedom*.

53 Betts and Collier, *Refugee*, ch. 6.

54 Terlazzo, "Conceptualizing Adaptive Preferences Respectfully."

there are secondary considerations excluding states or other agencies from the role of autonomy promoter and that in some cases there is most reason *not* to promote autonomy.

This specification of what people often have reasons to promote fits elegantly with a normative-capacity account. It does not fit well with formal accounts. When we think about the kinds of properties identified by formal accounts—reflective acceptance, ability to treat a consideration as a reason, answering for oneself in the social exchange—it is at least not obvious whether and why we have reasons to promote these things. Perhaps we do have reason to promote these things; much will depend on how the details are spelled out. But it surely is not obvious that we have quite general and powerful reasons to promote these properties. Contrast this with our confident commitment to promoting autonomy. Barring complications about special secondary reasons some agents might have not to be autonomy promoters, we think there are standing agent-neutral reasons to promote *anyone's* autonomy. This confidence is readily vindicated if autonomy turns out to require (i) valuable options and (ii) normative competence. We can readily appreciate why these twin goods would be valuable and worthy of promotion. Perhaps formal views can ultimately rise to the challenge of explaining why the properties they posit as constituents of autonomy are worthy of promotion. But the case needs to be made. There is at least a *prima facie* challenge here: normative-capacity accounts are well-positioned to make sense of the idea that we generally have reasons to promote people's autonomy; formal accounts, by contrast, are not so obviously well-positioned—whether they can make sense of the reasons we have to promote autonomy is more of an open question.

Perhaps, however, this is an unfair assessment of the situation. Consider the following problem. There is an ambiguity in the idea of reasons-responsiveness: Does it mean merely *having capacities* for responding to reasons or *actually exercising* those capacities? Which of these does the normative-capacity account of autonomy appeal to? Is the mere capacity for responding to reasons enough for autonomy or must one also exercise one's capacities in such a way as to conform to one's reasons? The label—normative-capacity account—certainly suggests the former. There are also systematic pressures encouraging this interpretation. Intuitively, it seems that choosing autonomously is not the same thing as choosing rightly or wisely or well.⁵⁵ And this intuition is underwritten by one of autonomy's central normative roles: if part of what makes paternalism presumptively wrong is that it conflicts with autonomy, it is hard to deny the possibility of autonomous bad choices; and if autonomous bad choices are

55 Cf. Raz, *The Morality of Freedom*, 412.

possible, autonomy cannot consist in appropriately exercising one's capacities for reasons-responsiveness. Moreover, the successful-exercise-of-normative-capacity interpretation of autonomy seems to imply that only the virtuous are really autonomous. If one wants a view of autonomy that squares with standard liberal commitments, the pure-capacity interpretation looks far more promising. But suppose the pure-capacity interpretation of autonomy is right. Then it is no longer so clear why autonomy is the sort of thing we generally have reasons to promote.⁵⁶ Merely having *capacities* for reasons-responsiveness, after all, is not all that valuable; what *is* valuable is having the capacities *and exercising them well*, i.e., actually succeeding in responding to one's reasons. The normative-capacity theorist cannot have her cake and eat it too: if she wants an account that makes sense of *one* of autonomy's key normative roles—being a bar to paternalism—she has to give up on being able to account for its *other* normative role—being the sort of thing we have reasons to promote.

There are a variety of ways the normative-capacity theorist might respond. Here I want to focus on two. The first is to deny the objection's presupposition and to insist that mere normative capacity *is* valuable and worth promoting after all. How so? Put briefly, normative capacity constitutes persons as responsible for their lives in a deep and meaningful sense—and *that* is good. To be sure, being responsible need not always and invariably be good: maybe being responsible for very bad decisions can make someone's life go worse. However, this qualification is consistent with the general and prospective value of autonomy as a thing worthy of promotion. Think of it this way. Persons are responsible agents. It is because of this fact that persons are capable of accessing or realizing special forms of value. Not all value in life is conditioned by being responsible, but some of it is. When autonomy is diminished, so is the opportunity for realizing these special forms of value. Hence, autonomy has prospective value as a generic life asset because it gives one the chance to live the most valuable kind of life. In general, and considered apart from what people make of their freedom, the interests of people who are robbed of the opportunity of being autonomous, and hence responsible, for their choices and lives are set back. The suggestion, in other words, is that we think of normative capacity as an *opportunity good*. Opportunities, of course, can be misused and wasted. But we often have reasons to provide people with valuable opportunities all the same. This vindicates, at least in a general way, the claim that we have reasons to promote normative capacity.

The second response accepts the objection's presupposition. It agrees that merely having normative capacities is not valuable and, hence, not the sort of thing we generally have reasons to promote, but insists that this only shows that

56 Thanks to David Brink and an anonymous reviewer from *JESP* for pressing this point.

we should reject a purely capacitarian account of autonomy. For the reasons I have given, this response may look unpromising. But I do not think this is necessarily so. Because I think we need to preserve the idea of autonomy as a bar to paternalism and hence the possibility of misguided autonomous choices, I think the response would be unpromising if it simply collapsed autonomy into a form of virtue without remainder. It need not, however, do that. T.H. Green distinguishes between two kinds of freedom: responsibility-entailing freedom and perfection-entailing freedom.⁵⁷ Putting this in terms of normative capacity, the former is the idea of having the ability to detect and pursue norms and values—reasons-responsiveness. This ability is plausibly at the basis of responsible agency, so the corresponding idea of freedom is a responsibility concept: one is responsible and can, in whatever way is suitable, be held responsible for one's choices. This ability can be had even if it lies dormant or is exercised poorly. The latter is the idea of realizing normative capacity—of actually successfully tracking relevant norms and values and then successfully *conforming* behavior accordingly. This ability is plausibly a kind of perfection of our rational natures, so the corresponding idea of freedom is a virtue concept: one realizes an important human excellence and merits approbation and esteem on that basis. Now one might plausibly hang on to *both* the responsibility concept and the virtue concept as essential to our thinking about autonomy. If that is right, then the second response can be put as follows: once we distinguish the responsibility concept from the virtue concept, we can see that what it is we have reasons to promote is the property corresponding to the virtue concept—perfection-entailing freedom. As long as we also accept that respecting responsibility-entailing freedom is important, we have not collapsed autonomy into virtue; instead, we have come to see that our thinking about autonomy is more complex than we might have initially thought.

I leave open which of these two responses is more compelling. Readers may find one or the other more compelling depending on how they think a normative-capacity account should be developed. For my part, I like them both. It seems to me that the first is partly right: there is *some* positive value to normative capacity as a generic opportunity good and this accounts for some of the reasons we have to promote autonomy. But it seems to me the second is partly right too: by itself, unexercised normative capacity is *not very* valuable; we also, and perhaps especially, have reasons to promote the fulfillment or appropriate exercise of normative capacity. This is not the place to develop a normative-capacity account in any detail, so I will just confine myself to the following brief and mostly suggestive remark. I think it is important to recognize different strands in our

57 Cf. Brink, *Perfectionism and the Common Good*, 81.

thinking about autonomy. Making use of the distinction introduced by Stephen Darwall between two forms of respect, we can recognize distinct normative statuses associated with our thinking about autonomy: recognition respect goes with being a responsible chooser, appraisal respect with how capacities are exercised.⁵⁸ Moreover, once we spell out autonomy's normative role, I think we will see that it commits us to both the capacity concept and the virtue concept, each one associated with a different status. A full and adequate specification of what we have reasons to promote will include the virtue notion; at the same time, we cannot dispense with the idea of negative autonomy rights attaching to responsible agency. The ambiguity between *mere capacity* and *fulfillment of capacity* is present in many of the normative-capacity accounts that have been offered in the literature. Once the ambiguity is noticed and the alternatives are clarified, there is an important intramural debate to be had about how best to develop such a view: Should we go for a pure capacity view or a pure virtue view? My sense is that we should go for neither: we should reject the choice as a false dichotomy and give up instead on the assumption that autonomy is a simple unitary thing. On closer inspection, I believe our thought about autonomy reveals two distinct ideals—the ideal of responsibility-entailing freedom and the ideal of perfection-entailing freedom—neither of which are dispensable. The challenge for such a hybrid account, of course, is to spell out the details in such a way that each strand of our thinking is part of a larger, complex whole, rather than merely two entirely separate and unrelated things. This is not the place to tackle that challenge. But I believe the hybrid option is worth investigating. For those who are suspicious of this option, believing that the virtue notion must be treated as a separate value and not confused with autonomy, but who nevertheless find formal views of autonomy wanting, I commend the pure capacity interpretation along with the first response above. If I'm right about the viability and attractiveness of the hybrid option, the pure capacity interpretation will turn out to be an impoverished conception of autonomy, though still superior to formal views.

Turn now to reasons of respect. Respecting a person's autonomy means at least two things. First, it means respecting the person's right to make self-regarding choices as she sees fit, including (perhaps up to some threshold) bad choices. This idea is, of course, closely associated with anti-paternalist norms in liberal societies. Second, it means honoring the person's perspective—their wishes, what matters to them, what they care about, and so on. The latter shows up, for exam-

58 Darwall, "Two Kinds of Respect."

ple, in what is required to treat someone of another religion with respect or (a bit more broadly) in respecting claims of conscience, whether religious or secular.⁵⁹

As we have seen, formal views of autonomy are commonly taken to be in the business of specifying conditions for authenticity. Suppose they succeed at this. Then it seems they have the ingredients for vindicating the second manifestation of respect, i.e., the one having to do with respect for conscience. Say you must decide whether to give a blood transfusion to an unconscious Jehovah's Witness to keep her alive. You know for a fact that she would not want to be given the blood transfusion, even though her life depends on it. It is not obvious what you ought, all things considered, to do. Still, whatever the right thing to do is, it seems there are powerful reasons of respect speaking in favor of honoring her (counterfactual) wishes not to receive the transfusion. Contrast this with a case where you know the religious commitments are superficial or have been inculcated in a suspect way. Perhaps the person has only been flirting with the Jehovah's Witness community for a couple of weeks or she has been drugged, manipulated, brainwashed, or coerced into having the commitments she does. In this case, presumably the weight that should be given to respecting the person's wishes is less, if any should be given to them at all. The difference between these cases is not in the content of the patient's request—that is the same. Instead, it is to be found in something like the position of the request vis-à-vis the person's authentic self.⁶⁰ Insofar as formal views are equipped to give us a story about authenticity, then, they are in a position to give us a story about this crucial dimension of respect for persons: honoring (i.e., giving some weight to) their point of view.

But it is not clear that formal views have the ingredients for vindicating the first manifestation of respect, i.e., the one having to do with the strong anti-paternalist presumption. As I argued above, authenticity plausibly suffices for attributability-responsibility, but this is not the right kind of responsibility to make sense of strong anti-paternalist practices. The strong liberal anti-paternalist presumption would seem best justified by the assumption that persons are more than merely attributability responsible for their choices. There are often good reasons *in favor* of paternalism, even when paternalism is all-things-considered wrong. In particular, people's welfare matters greatly and any balanced assessment of the ethics of paternalism must recognize this side of the balance sheet. How could a person's foolish choices merit protection? How could the kind of freedom that would allow people to make potentially ruinous life choices be justified? The mere fact that a choice is authentically an agent's own would not

59 Maclure and Taylor, *Secularism and Freedom of Conscience*.

60 Enoch, "Hypothetical Consent and the Value(s) of Autonomy."

seem to suffice to give other persons presumptively decisive reasons to desist from paternalistically interfering with the choice. A more robust form of responsibility seems to be required to make sense of that. This more robust form of responsibility would be secured by the ability to appreciate and appropriately respond to normative features relevant to choice.

To be clear, my point is about the *presuppositions* behind our *general* stance. I am not suggesting that considerations about normative capacity always feature, or always ought to feature, or that they are the only or always the most important considerations in every particular case in which the anti-paternalist presumption holds up. Nor am I claiming that there is not *some* anti-paternalist support from mere attributability-responsibility: there are, as I argued in the last paragraph, *pro tanto* reasons to honor people's points of view in self-regarding matters and these plausibly contribute to the case against paternalism. Rather, I am claiming that the kind of robust anti-paternalist norms characteristic of liberal social morality would be difficult to justify unless people were responsible (in the sense of being on the hook) for their choices. And this is difficult to make sense of in the absence of relevant normative capacities. Normative capacity is the ability to register and appropriately respond to normative features relevant to choice. Thus, if normative capacity is not required for autonomous choice, as formal accounts must maintain, this means one's choice about x can be autonomous, independent of any sensitivity to the features in virtue of which x is choice-worthy. But this is surely puzzling. For how can one be robustly and meaningfully responsible for choosing x in the absence of capacities for tracking what is relevant to the question of whether one should choose x ? The requirement that we honor people's perspective is not strong enough for quite general and robust anti-paternalist norms. Perhaps such considerations weigh in here and there in particular cases, but they cannot plausibly be thought to support a regime of vigorous anti-paternalist protections. Assuming that paternalism is (at least in part) presumptively wrong because it conflicts with people's sphere of "sovereignty," and assuming that autonomy-as-a-capacity is the ground of autonomy-as-a-right, we need to ask what view of autonomy capacities would be required to justify a robust sphere of self-sovereignty. Views of autonomy that include a normative-capacity condition would seem much better equipped than views that do not, to vindicate a robust sphere of self-sovereignty.

It is possible, of course, that people's *actual* normative capacities are often quite fragile. In that case, the idea that persons are normative agents, capable of tracking and responding to the normative features bearing on their choice, may be something of an idealization. One might maintain, as I suggested earlier, that the anti-paternalist norms are partly proleptic or educative, cultivating

the thing they appear to presuppose. Or one might maintain that, although it is something of an idealization, people are responsible often enough and, given the other contributing reasons against paternalism and perhaps secondary reasons against too closely tracking facts about normative competence, it is an acceptable idealization. My point is not to defend anti-paternalist norms but to clarify what we are plausibly committed to in accepting them. If someone thinks adults are *not* normatively competent *most* of the time, then it seems to me they should in principle be prepared to accept a much more invasive regime of paternalism than we tend to think appropriate in liberal social orders (even if in practice such a regime would be too difficult or too expensive or too unpopular or too abusive, etc.). For if adults are really not normatively competent most of the time, they will in this respect be a lot like children, and it will be difficult to see what principled objection would remain to treating them like children, except that there might be a variety of practical obstacles to doing so. If one thinks there are strong principled objections against paternalism—that the objections to it are not just incidental or technical—there is substantial pressure to also acknowledge that persons are, or can be, responsible agents. Even if our self-conception as responsible agents is slightly idealized, as long as it does not radically betray the facts about us, we can make sense of strong principled objections to paternalism. This self-conception of responsible agency—perhaps a mix of fact and ideal—is better captured by normative-capacity accounts than by formal accounts of personal autonomy.

In sum, autonomy is a recognizable value in liberal social orders, which prize self-direction and are committed to protecting a significant sphere for individual choice. At the heart of this social vision is the idea of persons as dignified choosers who must chart their own course through life.⁶¹ This idea marks out two normative roles for the idea of autonomy. One is an agency ideal. All else equal, autonomy is a desirable agency characteristic. Another is a principle protecting the exercise of autonomous agency. The choices of an autonomous agent call for respect. These are familiar ideas. And my argument in this section has been that, on the face of it, the normative role played by our concept of autonomy fits much more naturally with normative-capacity accounts of personal autonomy than formal accounts.

It is only fair to acknowledge that the situation has seemed to many philosophers to be exactly the reverse.⁶² Indeed, tension with perceived liberal commitments is a central source of resistance to normative-capacity accounts. For example, John Christman is a well-known defender of a formal view of autonomy, and

61 G. Dworkin, *The Theory and Practice of Autonomy*; Raz, *The Morality of Freedom*.

62 Thanks to an anonymous reviewer from *JESP* for pressing me on this.

he motivates the view in large part because of its coherence with what he takes to be the best interpretation of liberalism.⁶³ Normative-capacity accounts, argues Christman, are in tension with liberalism. They seem, among other things, to suggest a “sliding-scale” picture of anti-paternalist protections tailored to match the degree of decision-making competence and to invite state-sponsored perfectionist programs aimed at getting people’s choices to align with the true and the good. These concerns are serious. While I cannot vindicate the compatibility of normative-capacity accounts with liberalism, I wish to indicate at least briefly why I think such accounts cannot be too quickly dismissed as illiberal.

As I have already suggested, the idea of responsible agency seems a crucial bulwark in any principled anti-paternalist case, and normative-capacity accounts seem better positioned than formal accounts to interpret what this sort of responsible agency comes to. Moreover, normative-capacity accounts can, and I think should, accept the idea of negative autonomy rights. Most of us think (unjustified) paternalism *wrongs* people. We operate with the assumption that people have a *right* to decide for themselves in certain matters and that paternalism constitutes a usurpation of their *rightful authority* to do so.⁶⁴ A negative autonomy right protects a person (within suitable limits) in the making of bad self-regarding choices. This is just another way of saying that competent adults have powerful claims against others not to be interfered with in self-regarding matters. Since the right attaches to the capacity rather than its exercise, the right need not be thought of as conditional on making good choices.

One might, however, worry that even with negative autonomy rights in place, fidelity to the underlying normative structure would push normative-capacity accounts toward three unpalatable conclusions: (i) significant scope-restrictions on who enjoys negative autonomy rights, (ii) variations in autonomy levels and therefore different autonomy rights for different competent individuals, and (iii) an invitation to make minute discriminations among persons concerning their normative competence. But this is not necessarily so.

First, we should distinguish between scalar and threshold assessments of autonomy. On a plausible view, negative autonomy rights attach to threshold normative competence. It is a further question where to set the threshold, but there is no reason to suppose normative-capacity accounts are committed to setting it particularly high.⁶⁵ The lower the threshold is set, the less revisionary the account will be vis-à-vis standard liberal practice.

63 Christman, “Constructing the Inner Citadel,” “Liberalism and Individual Positive Freedom,” “Procedural Autonomy and Liberal Legitimacy,” and *The Politics of Persons*.

64 Darwall, “The Value of Autonomy and Autonomy of the Will,” 267–68.

65 Cf. Griffin, *On Human Rights*, 156; Kauppinen, “The Social Dimension of Autonomy,” 297.

Second, variation above the threshold does not necessarily yield differential allocation of rights. To be sure, there is an important question about how to resist this conclusion. But the problem is more general and is familiar from discussions of equality. Egalitarians are committed to ignoring variation above some threshold, treating persons as equals even when they exhibit morally relevant properties to different degrees.⁶⁶ Hence, the problem is no worse for normative-capacity theorists of personal autonomy than it is for egalitarians in general.

Third, it is open to normative-capacity theorists to say that above some threshold of competence, treating persons equally requires what Ian Carter calls “opacity respect,” that is, in such a way as to not make fine-grained distinctions about their normative competence.⁶⁷ If so, then there would be moral reasons to desist from too closely tracking or using information about normative competence above the threshold, at least by certain agencies and within specified contexts (e.g., the state in relation to its citizens).

Fourth, a normative-capacity account is consistent with thin procedurally defined operative standards in different domains. For example, one might think that a normative-capacity account would demand a stingy approach to medical consent, e.g., in who gets deemed “capacitous.” But this is not obvious. There may be good secondary practical and moral reasons for the existing standards, whether or not they adequately track normative competence. A variety of considerations—evaluative disagreement, proneness to error, liability to abuse, practical serviceability, and so on—speak in favor of thin procedural-looking proxy measures for normative competence that may, in practice, be overinclusive from the point of view of genuine normative competence. Since the pressures of crafting realistic and well-justified policy may license and even require a departure from attempting to use genuine normative competence as criterial for the determination of negative autonomy rights in various setting, we must be cautious about attributing to normative-capacity views pro-paternalist or illiberal policy implications in practice.

In short, normative-capacity accounts need not be wildly revisionary vis-à-vis widely accepted liberal views about equality, rights, and respect. But do they commit us to perfectionist politics? The answer, I think, is that they do not. It is one question what autonomy is; it is a further question how autonomy is to be promoted—and by whom. Even if there are quite general agent-neutral reasons to promote anyone’s autonomy, there may be good secondary reasons for insisting that it is not everyone’s business to promote everyone else’s autonomy, and in

66 Cf. Waldron’s discussion of equality in terms of the idea (originally from Rawls) of a “range property” (*One Another’s Equals*, 84–127).

67 Carter, “Respect and the Basis of Equality.”

particular, there may be special reasons to insist that states not be in the business of promoting autonomy. Whether the state may promote its citizens' autonomy is an important question, but it is orthogonal to the nature of autonomy. To see this, notice that it arises whether one adopts a normative-capacity view or a formal view. Suppose, for example, that autonomy is, as Gerald Dworkin maintains, "a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values."⁶⁸ This type of view, too, might be combined with either perfectionist or anti-perfectionist political commitments. There is nothing about formal views of autonomy that commits those who hold such views to say to states or other agencies in authority, "Hands off on promoting this property!" A formal theorist might welcome state intervention in promoting autonomy, e.g., by promoting critical reflection. Conversely, there is nothing about normative-capacity accounts that commits those who hold such views to say, "This property may (or should) be promoted by the state." Whether one has an inviting posture to state intervention is orthogonal to which view one takes about the nature of autonomy. The debate between liberal perfectionists and liberal anti-perfectionists—interesting and important though it is—should not drive our theorizing about personal autonomy.

Suppose, however, that normative-capacity accounts do invite perfectionism in politics. Would this really be damning news? I think that is far from obvious. Some philosophers take it as virtually axiomatic for an account of autonomy of a liberal bent that it must respect neutrality and safeguard anti-perfectionism in politics.⁶⁹ But the question of how best to interpret the requirement of state neutrality is notoriously complex and controversial. Proponents of formal theories all too often simply take for granted that liberalism favors their view. Yet liberalism is a broad camp. There are sensible forms of perfectionist liberalism that have as good a claim as Rawlsian justificatory liberalism to being *bona fide* versions of liberalism.⁷⁰ To suggest that *all* substantive accounts of autonomy are illiberal will not work: normative-capacity accounts, as I have suggested, need not have radically illiberal implications—and they do a good job of interpreting the picture of responsible agency that seems to be presupposed by our liberal

68 G. Dworkin, *The Theory and Practice of Autonomy*, 20.

69 Cf. Christman, *The Politics of Persons*; G. Dworkin, *The Theory and Practice of Autonomy*; Westlund, "Rethinking Relational Autonomy."

70 E.g., Green, *Lectures on the Principles of Political Obligation and Other Writings*; Hurka, *Perfectionism*; Mill, *On Liberty*; Raz, *The Morality of Freedom*; Sher, *Beyond Neutrality*; and Wall, *Liberalism, Perfectionism, and Restraint*. On Mill as a perfectionist liberal, see Brink, *Mill's Progressive Principles*.

anti-paternalist practice. There is work to be done interpreting liberalism. At the very least, I think those who leverage anti-perfectionist arguments in favor of formal accounts of autonomy have more work to do in showing why we should antecedently favor non-perfectionist over perfectionist liberalism. And even if they make this case convincingly, it does not, so far as I can see, follow that autonomy is best interpreted in formal terms. For as I suggested in the last paragraph, what autonomy is and who gets to promote it are separate questions.

I conclude that normative-capacity accounts need not necessarily conflict with liberal commitments. Much more, of course, would need to be said to allay fears that normative-capacity accounts commit us to unattractive views of politics. What I have tried to show here, at least in brief outline, is that the conflict between liberalism and at least one variant of a substantive view of autonomy may not be as sharp or deep as sometimes supposed.

5. CONCLUSION

I have argued that formal views of autonomy face serious challenges. In particular, I have argued that they do not give us the right building blocks to make sense of the kind of responsibility that is plausibly at stake in autonomy, that they posit a fact/value asymmetry that creates an explanatory burden, and that they supply rather meager resources for helping us make sense of autonomy's normative role. Moreover, I have suggested that normative-capacity accounts need not be on a collision course with liberal commitments. None of this is decisive. Perhaps formal accounts can be developed in such a way that they meet the *prima facie* challenges I have sketched. And perhaps normative-capacity accounts are, after all, in significant tension with liberal commitments—or face other insurmountable problems. My aim in this paper has not been to give a full-scale assessment of competing theories, but to offer some reasons for rethinking the status quo. The choice between formal and normative-capacity accounts of personal autonomy is deep and consequential. It represents a quite fundamental fork in the road that any theorist of personal autonomy must face. Many philosophers have bounded down the formal path, thinking it would take them in the right direction. But if the arguments in this paper are along the right lines, it may be time to revisit the fork in the road and consider going the other way.⁷¹

University of Graz
jonathan.knutzen@gmail.com

71 Thanks to Richard Arneson, David Brink, Dana Nelkin, Manuel Vargas, and two anonymous reviewers from *JESP* for helpful input on an earlier version of this paper.

REFERENCES

- Arneson, Richard J. "Mill versus Paternalism." *Ethics* 90, no. 4 (July 1980): 470–89.
- Arpaly, Nomy. "Responsibility, Applied Ethics, and Complex Autonomy Theories." In Taylor, *Personal Autonomy*, 162–81.
- . *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford: Oxford University Press, 2003.
- Benn, S. I. "Freedom, Autonomy and the Concept of a Person." *Proceedings of the Aristotelian Society* 76 (1975–76): 109–30.
- Benson, Paul. "Feminist Intuitions and the Normative Substance of Autonomy." In Taylor, *Personal Autonomy*, 124–42.
- . "Feminist Second Thoughts about Free Agency." *Hypatia* 5, no. 3 (September 1990): 47–64.
- . "Free Agency and Self-Worth." *Journal of Philosophy* 91, no. 12 (December 1994): 650–68.
- . "Freedom and Value." *Journal of Philosophy* 84, no. 9 (September 1987): 465–86.
- Berofsky, Bernard. *Liberation from Self: A Theory of Personal Autonomy*. Cambridge: Cambridge University Press, 1995.
- Betts, Alexander, and Paul Collier. *Refuge: Rethinking Refugee Policy in a Changing World*. Oxford: Oxford University Press, 2017.
- Bratman, Michael E. "Intention, Practical Rationality, and Self-Governance." *Ethics* 119, no. 3 (April 2009): 411–43.
- Brink, David O. *Mill's Progressive Principles*. Oxford: Oxford University Press, 2013.
- . *Perfectionism and the Common Good: Themes in the Philosophy of T.H. Green*. Oxford: Clarendon Press, 2003.
- Brink, David O., and Dana K. Nelkin. "Fairness and the Architecture of Responsibility." In *Oxford Studies in Agency and Responsibility*, vol. 1, edited by David Shoemaker, 284–314. Oxford: Oxford University Press, 2013.
- Buss, Sarah. "Autonomous Action: Self-Determination in the Passive Mode." *Ethics* 122, no. 4 (July 2012): 647–91.
- Carter, Ian. "Respect and the Basis of Equality." *Ethics* 121, no. 3 (April 2011): 538–71.
- Christman, John. "Autonomy in Moral and Political Philosophy." *Stanford Encyclopedia of Philosophy* (Spring 2018). <https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/>.

- . “Autonomy and Personal History.” *Canadian Journal of Philosophy* 21, no. 1 (March 1991): 1–24.
- . “Constructing the Inner Citadel: Recent Work on the Concept of Autonomy.” *Ethics* 99, no. 1 (October 1988): 109–24.
- . “Liberalism and Individual Positive Freedom.” *Ethics* 101, no. 2 (January 1991): 343–59.
- . *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge: Cambridge University Press, 2009.
- . “Procedural Autonomy and Liberal Legitimacy.” In Taylor, *Personal Autonomy*, 277–98.
- Christman, John, and Joel Anderson. “Introduction.” In *Autonomy and the Challenges to Liberalism: New Essays*, 1–26. Cambridge: Cambridge University Press, 2005.
- Darwall, Stephen. “Two Kinds of Respect.” *Ethics* 88, no. 1 (October 1977): 36–49.
- . “The Value of Autonomy and Autonomy of the Will.” *Ethics* 116, no. 2 (January 2006), 263–84.
- Dworkin, Gerald. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press, 1988.
- Dworkin, Ronald. *Life’s Dominion: An Argument about Abortion, Euthanasia, and Individual Freedom*. New York: Vintage Press, 1993.
- Ekstrom, Laura Waddell. “Autonomy and Personal Integration.” In Taylor, *Personal Autonomy*, 143–61.
- Enoch, David. “Hypothetical Consent and the Value(s) of Autonomy.” *Ethics* 128, no. 1 (October 2017): 6–36.
- Feinberg, Joel. *The Moral Limits of the Criminal Law*. Vol. 3, *Harm to Self*. Oxford: Oxford University Press, 1986.
- Fischer, John Martin, and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, 1998.
- Frankfurt, Harry G. “Autonomy, Necessity, and Love.” In *Necessity, Volition, and Love*, 129–41. Cambridge: Cambridge University Press, 1999.
- . “Freedom of the Will and the Concept of a Person.” *Journal of Philosophy* 68, no. 1 (January 1971): 829–39.
- Friedman, Marilyn. *Autonomy, Gender, Politics*. Oxford: Oxford University Press, 2003.
- Green, T. H. *Lectures on the Principles of Political Obligation and Other Writings*, edited by Paul Harris and John Morrow. Cambridge: Cambridge University Press, 1986.
- Griffin, James. *On Human Rights*. Oxford: Oxford University Press, 2008.

- Groll, Daniel. "Paternalism, Respect, and the Will." *Ethics* 122, no. 4 (July 2012): 692–720.
- Hurka, Thomas. *Perfectionism*. Oxford: Oxford University Press, 1993.
- . "Why Value Autonomy?" *Social Theory and Practice* 13, no. 3 (Fall 1987): 361–82.
- Kauppinen, Antti. "The Social Dimension of Autonomy." In *Axel Honneth: Critical Essays*, edited by Danielle Petherbridge, 255–302. Leiden, Netherlands: Brill, 2011.
- Killmister, Suzy. "Autonomy and False Beliefs." *Philosophical Studies* 164, no. 2 (June 2013), 513–31.
- . *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. New York: Routledge, 2018.
- Mackenzie, Catriona. "Three Dimensions of Autonomy: A Relational Analysis." In *Autonomy, Oppression, and Gender*, edited by Andrea Veltman and Mark Piper, 15–41. Oxford: Oxford University Press, 2014.
- Mackenzie, Catriona, and Natalie Stoljar. "Introduction: Autonomy Refigured." In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, 3–34. Oxford: Oxford University Press, 2000.
- Maclure, Jocelyn, and Charles Taylor. *Secularism and Freedom of Conscience*. Cambridge, MA: Harvard University Press, 2011.
- McDowell, John. "Autonomy and Its Burdens." *Harvard Review of Philosophy* 17, no. 1 (2010): 4–15.
- McKenna, Michael. "The Relationship Between Autonomous and Morally Responsible Agency." In Taylor, *Personal Autonomy*, 205–34.
- Mele, Alfred. *Autonomous Agents: From Self Control to Autonomy*. Oxford: Oxford University Press, 1995.
- Meyers, Diana Tietjens. *Being Yourself: Essays on Identity, Action, and Social Life*. Lanham, MA: Rowman and Littlefield, 2004.
- . "Decentralizing Autonomy: Five Faces of Selfhood." In *Autonomy and the Challenges to Liberalism: New Essays*, 27–55. Cambridge: Cambridge University Press, 2005.
- Mill, John Stuart. *On Liberty*. Edited by George Kateb and David Bromwich. New Haven: Yale University Press, 2003.
- Nelkin, Dana Kay. *Making Sense of Freedom and Responsibility*. Oxford: Oxford University Press, 2011.
- Nussbaum, Martha C. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: Harvard University Press, 2011.
- Oshana, Marina A.L. "Personal Autonomy and Society." *Journal of Social Philosophy* 29, no. 1 (March 1998): 81–102.

- . *Personal Autonomy in Society*. New York: Routledge, 2006.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Oxford University Press, 1986.
- Rosen, Gideon. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103, no. 1 (2003): 61–84.
- Savulescu, Julian. "Rational Non-interventional Paternalism: Why Doctors Ought to Make Judgments of What Is Best for Their Patients." *Journal of Medical Ethics* 21, no. 6 (1995): 327–31.
- Sayre-McCord, Geoffrey, and Michael Smith. "Desires . . . and Beliefs . . . of One's Own." In *Rational and Social Agency: The Philosophy of Michael Bratman*, edited by Manuel Vargas and Gideon Yaffe, 129–51. Oxford: Oxford University Press, 2014.
- Sen, Amartya. *Development as Freedom*. Oxford: Oxford University Press, 1999.
- Sher, George. *Beyond Neutrality: Perfectionism and Politics*. Cambridge: Cambridge University Press, 1997.
- Stoljar, Natalie. "Autonomy and the Feminist Intuition." In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, edited by Catriona Mackenzie and Natalie Stoljar, 94–111. Oxford: Oxford University Press, 2000.
- . "Feminist Perspectives on Autonomy." *Stanford Encyclopedia of Philosophy* (Winter 2018). <https://plato.stanford.edu/archives/win2018/entries/feminism-autonomy/>.
- Taylor, James Stacey, ed. *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press, 2005.
- Terlazzo, Rosa. "Conceptualizing Adaptive Preferences Respectfully: An Indirectly Substantive Account." *Journal of Political Philosophy* 24, no. 2 (June 2016): 206–26.
- Vargas, Manuel. *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press, 2013.
- Waldron, Jeremy. *One Another's Equals: The Basis of Human Equality*. Cambridge, MA: Harvard University Press, 2017.
- Wall, Steven. *Liberalism, Perfectionism, and Restraint*. Cambridge: Cambridge University Press, 1998.
- Watson, Gary. "Two Faces of Responsibility." *Philosophical Topics* 24, no. 2 (Fall 1996): 227–48.
- Westlund, Andrea C. "Rethinking Relational Autonomy." *Hypatia* 24, no. 4 (Fall 2009): 26–49.
- Wolf, Susan. *Freedom within Reason*. Oxford: Oxford University Press, 1994.