

## KANTIAN PERSONAL AUTONOMY

ROBERT S. TAYLOR  
*University of California–Davis*

*Jeremy Waldron has recently raised the question of whether there is anything approximating the creative self-authorship of personal autonomy in the writings of Immanuel Kant. After considering the possibility that Kantian prudential reasoning might serve as a conception of personal autonomy, I argue that the elements of a more suitable conception can be found in Kant's Tugendlehre, or "Doctrine of Virtue"—specifically, in the imperfect duties of self-perfection and the practical love of others. This discovery is important for at least three reasons: first, it elucidates the relationship among the various conceptions of autonomy employed by personal-autonomy theorists and contemporary Kantians; second, it brings to the surface previously unnoticed or undernoticed features of Kant's moral theory; and third, it provides an essential line of defense against certain critiques of contemporary Kantian theories, especially that of John Rawls.*

**Keywords:** *autonomy; prudence; virtue; Kant; Rawls*

**T**he contemporary ideal of personal autonomy is one that celebrates creative self-authorship and encourages the development of those virtues (such as integrity and authenticity) that support it. It has its roots in the Romantic liberalism of John Stuart Mill, who argued for a multiplicity of “experiments in living” for the sake of the “free development of individuality.”<sup>1</sup> Simultaneously, however, it emphasizes the need for self-control to ensure that lower-order desires do not usurp the authority of the higher-order ones that define our character and determine the shape and direction of our lives. In sum, it weds the freewheeling creativity of the bohemian to the disciplined management of the bourgeois.

---

*AUTHOR'S NOTE: I wish to thank Robert Adcock, Yvonne Chiu, Meir Dan-Cohen, Brad Inwood, Chris Kutz, David Lieberman, Eric Rakowski, Connie Rosati, Sam Scheffler, Lael Weis, Allen Wood, and the other participants in workshops at Stanford and the University of California–Berkeley for their helpful comments. I am especially grateful to John Christman for his many excellent written suggestions; the essay is much improved because of them.*

POLITICAL THEORY, Vol. 33 No. 5, October 2005 602-628

DOI: 10.1177/0090591705278397

© 2005 Sage Publications

Jeremy Waldron has recently raised the question of whether there is anything approximating personal autonomy in the writings of Immanuel Kant.<sup>2</sup> Although Waldron ultimately concludes that Kant has *at most* a prototheory of personal autonomy, even so cautious a judgment would be considered controversial by both personal-autonomy theorists and contemporary Kantians. These two groups believe they use very different, even incompatible conceptions of autonomy and are deeply suspicious of the idea that these conceptions may be connected in fundamental ways. For example, Joseph Raz, a personal-autonomy theorist, warns that “personal autonomy should not be confused with the only very indirectly related notion of moral autonomy” and that Kantian moral autonomy in particular reduces “[self-]authorship to a vanishing point as it allows only one set of principles which people can rationally legislate and they are the same for all.”<sup>3</sup> Similarly, Onora O’Neill, a contemporary Kantian, thinks that Kant and personal-autonomy theorists have “radically differing conceptions of action and autonomy” and makes the case for “disentangling Kant’s conception of autonomy from contemporary ones.”<sup>4</sup> In this essay, I intend to further entangle these two conceptions—and perhaps initiate a process of reconciliation between these two groups of theorists—by arguing for a specifically *Kantian* personal autonomy.

A conception of personal autonomy could be Kantian in at least two different senses, though. First, it could be Kantian in the simple sense that it is present in Kant’s writings. I will contend, for example, that Kant’s idea of prudential reasoning is a conception of personal autonomy that would (or should) be recognized as such by contemporary personal-autonomy theorists. Second, it could be Kantian in the arguably deeper sense that Kant *himself* would recognize it as a species of autonomy. I will argue that prudential reasoning is *not* Kantian in this second sense but that one can find the elements of such a conception in Kant’s *Tugendlehre*, or “Doctrine of Virtue”—to wit, in the imperfect duties of self-perfection and the practical love of others. Only this second, more authentic version of Kantian personal autonomy can call into question the sharp contrast between personal and moral autonomy drawn by personal-autonomy theorists and contemporary Kantians alike and lay the groundwork for a resolution of their disagreement(s).

Locating twentieth-century concepts in the writings of eighteenth-century thinkers will strike some readers as a peculiar pastime if not a hopelessly anachronistic one, but I think this exercise has great value for at least three different reasons. First, it elucidates the relationship among the various conceptions of autonomy employed by personal-autonomy theorists and contemporary Kantians. As Gerald Dworkin notes, we face “one concept and many conceptions of autonomy,” with disagreement even among personal-autonomy theorists as to its meaning.<sup>5</sup> By exploring the question of whether

Kant had a conception of personal autonomy, we can learn much about the way these conceptions are (or are not) connected, thereby making some initial progress toward a unified theory of autonomy.<sup>6</sup>

Second, we can discover a great deal about Kant's own moral theory by confronting it with contemporary concepts and concerns. In the present case, by examining both the similarities and differences between personal autonomy and certain Kantian concepts (e.g., prudential reasoning and imperfect duties of virtue), we can bring to the surface previously unnoticed or undernoticed features of his theory (e.g., his concern for self-authorship). The danger, of course, is in reading concepts back into his theory that are in fact rather alien to it, but I think this danger is more than compensated for by the fresh insights that can potentially be gained by making the comparisons; moreover, this danger can be mitigated by a cautious and sensitive reading of the primary text.

Finally, the second, preferred version of Kantian personal autonomy offers an essential line of defense against certain critiques of contemporary Kantian theories, especially that of John Rawls. Jeremy Waldron and Michael Sandel, for example, have made two criticisms of Rawlsian agency (what I will call the "Discontinuity" and "Arbitrariness" criticisms, respectively), criticisms that can be deflected by Rawls's adoption of Kantian personal autonomy or of something close to it. Whether such an adoption can take place after his so-called "political turn" is another question entirely, to be considered below.

The remainder of my essay will be organized as follows. First, I will offer a very brief review of the literature of personal autonomy in order to orient the unfamiliar reader. Next, following a path partially blazed by Waldron, I will examine the remarkable similarities between personal autonomy and Kant's idea of prudential reasoning. Then, after arguing that Kant himself would *not* consider prudential reasoning to be a species of autonomy, I will discover the elements of a more suitable conception of personal autonomy in Kant's imperfect duties of self-perfection and the practical love of others. Finally, I will show how such a conception, if adopted by Rawls, can provide an effective line of defense against the criticisms mentioned above—though only at the perhaps modest cost of returning to the comprehensive Kantian liberalism of *Theory*.

#### *PERSONAL AUTONOMY: A BRIEF OVERVIEW*

Both Joseph Raz and John Rawls provide good general descriptions of personal autonomy in their respective writings. According to Raz:

The ideal of personal autonomy . . . holds the free choice of goals and relations as an essential ingredient of individual well-being. The ruling idea behind the ideal of personal autonomy is that people should make their own lives. The autonomous person is (part) author of his own life. The ideal of personal autonomy is the vision of people controlling, to some degree, their own destiny, fashioning it through successive decisions throughout their lives.<sup>7</sup>

Raz associates this ideal of self-authorship or self-creation with certain auxiliary virtues, such as self-awareness and integrity, and contrasts it with self-realization and other substantive ideals.<sup>8</sup> On Raz's understanding, what makes a person autonomous is not the specific content of his "goals and relations" but rather the manner in which he evaluates, adopts, and pursues them.

Rawls develops his own conception of personal autonomy (which he refers to as "rational autonomy"), one which places greater emphasis on its planning and deliberative features. He associates this kind of autonomy with Kantian empirical practical reason, saying at one point that it "roughly parallels Kant's notion of hypothetical imperatives."<sup>9</sup> Rather than focusing on the creative side of personal autonomy, Rawls highlights its dependence on principles of rational choice, such as "the adoption of effective means to ends; the balancing of final ends by their significance in our plan of life as a whole; and finally, the assigning of a greater weight to the more likely consequences."<sup>10</sup> For Rawls, personal autonomy is firmly wedded to deliberative rationality; as with Raz, though, the focus is on the *process* of deliberation rather than its outcome.

Some writers, including Gerald Dworkin and (indirectly) Harry Frankfurt, have argued that personal autonomy must involve a particular kind of preference structure as well as a particular kind of deliberation process. Dworkin, for example, defines personal autonomy as "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," where "higher-order" preferences "define their nature, give meaning and coherence to their lives."<sup>11</sup> Such a hierarchical preference structure may already be implicit in Rawls's theory, as he distinguishes plans of life from lower-order desires.<sup>12</sup>

As noted by Raz, personal autonomy is associated with a number of virtues, seen variously as conditions of its exercise or as constitutive elements. John Christman mentions the conditions of competency (e.g., rationality) and authenticity (reflective endorsement of one's desires).<sup>13</sup> Joel Feinberg offers a more comprehensive list, including authenticity, self-creation, integrity, and initiative.<sup>14</sup> Various types of self-control (over actions, belief, willing, etc.) are also relevant, as Alfred Mele points out, because *akrasia* makes it difficult to devise and carry out plans of life.<sup>15</sup> These diverse lists indicate

the underlying complexity and ambiguity of the idea of personal autonomy; it is a concept with many conceptions, as Dworkin aptly notes, and different theorists will understand it in different and even incompatible ways.<sup>16</sup>

Theorists of personal autonomy appear to be in agreement about one of its features, however: its distinctness from *moral* autonomy. Whether moral autonomy is understood in its specifically Kantian sense or more generally as the capacity for a sense of justice, these theorists believe that it neither implies nor is implied by personal autonomy. David Johnston, for example, argues that “a person could be morally autonomous without being in the habit of subjecting her own values and projects to critical appraisal. Similarly . . . it is possible to be personally autonomous without being morally autonomous,” as would be the case with a self-authoring, self-aware, and authentic mass murderer; the pure proceduralism of personal autonomy does not assure results consistent with the moral law or any other substantive standard.<sup>17</sup> As we shall see, this disjuncture between personal and moral autonomy is the main obstacle to constructing a genuinely *Kantian* personal autonomy—though not an insurmountable one.

#### *KANTIAN PRUDENCE AS PERSONAL AUTONOMY*

Jeremy Waldron asks “whether there is anything approximating personal autonomy in Kant’s account of happiness” and offers a three-page sketch of an affirmative answer, but he ultimately concludes that the passages he cites “do not add up to a theory of personal autonomy in the sense used by modern liberals,” though “they point a little bit in that direction.”<sup>18</sup> I will argue in this section that the parallels between Kantian prudential reasoning and personal autonomy are much stronger than Waldron believes and that this capacity for prudence is even worthy of respect—at least according to several respected secondary sources.

One task of Kantian prudential reasoning is to unite or integrate the inclinations into a single scheme of happiness. Kant says that in the idea of happiness “all inclinations unite in one sum” and that the function of prudence is to “unite all [one’s] purposes to [one’s] own enduring advantage.”<sup>19</sup> His language here sounds hedonistic, as if the function of prudence were merely to maximize a sum of pleasures, but in later writings it is less so, as when he says that inclinations “can be brought into a tolerable system” and that we must “curb them, so that they will not wear each other out but will instead be harmonized into a whole called happiness.”<sup>20</sup> Here, inclinations are not united under a single hedonistic metric but rather harmonized, systematized, and curbed where necessary; the heterogeneity of inclinations is maintained, and

prudential reasoning takes on more of an artistic than a mechanical quality.<sup>21</sup> Rawls uses nearly identical language when he describes the construction of a plan of life, one of personal autonomy's primary tasks:

The aim of deliberation is to find that plan which best organizes our activities and influences the formation of our subsequent wants so that our aims and interests can be fruitfully combined into one scheme of conduct. Desires that tend to interfere with other ends, or which undermine the capacity for other activities, are weeded out; whereas those that are enjoyable in themselves and support other aims as well are encouraged.<sup>22</sup>

Thus, in both Kantian prudential reasoning and the modern literature of personal autonomy, the integration of inclinations through a plan of life requires a judicious blend of balancing, pruning, and cultivating, a process closer to gardening than to mechanics.<sup>23</sup>

The inclinations to be integrated by Kantian prudential reason are not merely heterogeneous but also hierarchically related. Kant's example of the man suffering from gout who sacrifices the long-term happiness that comes from good health to the temporary pleasures of drink illustrates this point: the inclination to good health is, Kant here implies, more essential to happiness than the inclination to drink, and the latter should give way to the former on grounds of prudence.<sup>24</sup> He notes on another occasion that:

We can find satisfaction in the mere *exercise of our powers*, in the consciousness of our strength of soul in overcoming obstacles opposed to our plans, in cultivating our talents of spirit, and so forth, and we correctly call these joys and delights *more refined* because they are more under our control than others, do not wear out but rather strengthen feeling for further enjoyment of them, and while they delight they at the same time cultivate.<sup>25</sup>

As we have seen, such a hierarchical preference structure, with higher- and lower-order desires, is a key feature of personal autonomy as developed in the writings of Dworkin and Frankfurt.

Kant emphasizes that the process of integrating our inclinations in the pursuit of happiness is a task for *empirical* practical reason, which helps us harmonize desires of different kinds and of different orders of precedence. Reason cannot issue determinate principles for the pursuit of happiness, however, but only prudential *counsels*, such as "regimen, frugality, courtesy, reserve, and so forth, which experience teaches are most conducive to well-being on the average."<sup>26</sup> The successful pursuit of happiness requires life experience plus the judgment and discernment that can grow with it, so that we may learn in time what makes *us* happy as individuals. The counsels of prudence, Kant says, are "so tenuous that everyone must be allowed countless exceptions in order to adapt his choice of a way of life to his particular

inclinations and his susceptibility to satisfaction and still, in the end, to become prudent only from his own or others' misfortunes."<sup>27</sup> Kantian prudential reasoning thus combines the idiosyncratic self-authorship emphasized by Raz with the deliberative rationality stressed by Rawls.

As noted in the last section, personal autonomy is associated with a number of virtues, such as authenticity, integrity, and so on. Although Kant usually discusses these virtues in a *moral* context (e.g., self-awareness or self-knowledge), they also show up in his analysis of prudence.<sup>28</sup> Self-control, for example, is mentioned as a quality of character that is "good for all sorts of purposes [and] even seem[s] to constitute a part of the *inner* worth of a person" but is not unconditionally good because it can support prudent but immoral behavior.<sup>29</sup> Self-control's complementary vice, *akrasia*, appears to be responsible for the failure of prudence in the gout case mentioned above. Paul Guyer believes this case exemplifies the truth that happiness is "a rational goal that must be represented as distinct from and authoritative over one's momentary inclinations"; self-control is precisely the virtue that can render the authority of prudential reasoning efficacious.<sup>30</sup>

Most theorists of personal autonomy believe it to be, as Raz does, "an essential ingredient of individual well-being," and Kant similarly holds that the happiness achieved through the exercise of prudence is a good in itself (albeit not an unconditional one). Happiness, for example, is part of the highest good in Kant's moral theory: virtue is the *supreme* good, but the *complete* good requires happiness proportional to virtue as well, at least for sentient beings such as ourselves.<sup>31</sup> Moreover, in his later writings Kant argues that our original predispositions to self-love, which have happiness as their object, are predispositions to the good, and therefore any attempt to root them out would "not only be futile but harmful and blameworthy as well."<sup>32</sup>

Some theorists of personal autonomy have made even stronger claims in its favor: Gerald Dworkin, for instance, argues that "moral respect is owed to all because all have the capacity for defining themselves."<sup>33</sup> Kant would seem unlikely to make similar claims in favor of prudential reasoning, as respect in Kant is usually directed toward man as a *moral* agent, capable of acting consistently with and out of respect for the moral law, rather than as a prudential agent. Several commentators on Kant, however, maintain that there is a Kantian argument for respecting man as a *prudential* agent. Henry Allison, for example, argues that our freedom and rational agency are exemplified by the pursuit of happiness as well as morality: *all* maxims, including maxims of prudence and skill, are freely adopted (i.e., objects of choice or *Willkür*) and also "subject to the objective criteria of reasonableness expressed in the objective practical principles," such as the principle expressed in the *Groundwork* that "whoever wills the end also wills . . . the indispensably necessary

means to it that are within his power.”<sup>34</sup> As Christine Korsgaard points out, Kant says that “the capacity to set an end for oneself—any end whatsoever [moral or nonmoral]—is what characterizes humanity”; this claim, when combined with the Formula of Humanity as an End-in-Itself (“so act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means”), suggests that even our capacity for *prudential* end-setting is worthy of respect.<sup>35</sup> Finally, Allen Wood argues that our self-conception and even sense of self-worth derive in part from our capacity for prudence:

The difference between acting on technical imperatives and acting on prudential imperatives (or counsels of prudence) may be seen as a difference in the depth of the self-conception that gives these practical principles their authority. What binds me to act on a technical imperative is the conception of myself as a rational being who happens to have set a certain end. What binds me to act on a counsel of prudence is something still deeper, the conception of myself as a single self with a conception of its own good. My commitment to an end I have set can be abandoned at my discretion, but my rational commitment to prudence cannot be abandoned as if it were an arbitrary end.<sup>36</sup>

#### THE HETERONOMY OF KANTIAN PRUDENCE

The previous section provides compelling evidence, I believe, that prudential reasoning is a Kantian personal autonomy, at least in the *first* sense of the term: it is a conception of personal autonomy, recognizable as such by personal-autonomy theorists, that is found in Kant’s writings. In order for it to be Kantian in the second, deeper sense, however, Kant himself would need to be able to recognize it as a species of autonomy; without such recognition, the gap between personal and moral autonomy would remain unbridged, and we would merely have identified two parallel but ultimately nonintersecting accounts of autonomy in Kant’s practical philosophy. As I shall argue, though, Kant would *not* characterize prudential reasoning as autonomous, for two separate but closely related reasons. First, our desires are disorderly and constantly shifting; consequently, our conception of our own happiness, which tries to unite these desires into one coherent system, will be similarly fluctuating and perhaps *anomic*. Second, even if prudence could successfully systematize our desires, the *source* of these desires would still be outside the self, in nature and in corrupt society; therefore, even the orderly fulfillment of them would not be autonomous but rather *heteronomous*. Let us examine these two objections in turn.

Kant consistently describes our inclinations, which are the motivational elements of our plans of happiness, as inconstant and unpredictable. The

cause of this variability, according to Kant, is that “all the elements that belong to the concept of happiness are without exception empirical, that is, they must be borrowed from experience”; we are therefore uncertain about not only our current inclinations but also our future ones, which may be influenced by both our own actions and the actions of others.<sup>37</sup> Any plan of happiness built on such uncertain foundations will itself be uncertain, which is why Kant calls it a “fluctuating idea” and “not an ideal of reason but of imagination.”<sup>38</sup> As the task of prudence is the construction and implementation of such a plan of happiness, prudential reasoning will strive without success to systematize what is intrinsically unsystematic, even anomic or lawless. A necessary (but not sufficient) condition of autonomy is lawfulness; therefore, prudential reasoning cannot be characterized as autonomous. Just as J. B. Schneewind reminds us:

The natural disarray of the passions and desires is a given in Kant’s ethics. . . . The material of our lives, the substantive happiness we pursue, comes from forces we can neither create nor (as Kant eventually came to admit) eradicate. Moral agency brings a kind of unity to our lives that prudential agency alone could never bring.<sup>39</sup>

Suppose for a moment, however, that Kant is wrong about this and that prudential reasoning *is* capable of imposing a lawful plan of happiness on our disorderly inclinations. One might even suppose (as Kant does for the sake of argument) complete unanimity among finite rational beings regarding a plan of happiness, with respect to both ends and means.<sup>40</sup> Would this degree of order be sufficient to make prudential reasoning autonomous? Hardly, for autonomy requires not just lawfulness but a *self-originating* lawfulness, and Kant held throughout his life that inclinations originate *outside* the self, so that even the orderly fulfillment of them would be heteronomous.

According to Kant, *all* inclinations ultimately originate in nature: “happiness contains all (*and also not more than*) that which nature provides us.”<sup>41</sup> Those inclinations, imposed from without, are “always burdensome to a rational being” and make him yearn for “a deliverance from the manifold dissatisfaction in which all those needs entangle him.”<sup>42</sup> He experiences the pull of nature in him as an alien, even a degrading force, a “commission from the side of his sensibility which [he] cannot refuse”; without the “higher purpose” offered by the idea of morality he would remain trapped by nature in “mere animality.”<sup>43</sup> At times like these, Kant’s practical philosophy veers dangerously close to Manichaeism.<sup>44</sup>

As we saw in the last section, however, Kant appears to take a different position in his later writings, where he speaks of our original predispositions to self-love, which have happiness as their object, as predispositions to the

good. The *tone* of the later writings is certainly different, but they do not in fact differ substantively from the earlier writings. First, Kant continues to maintain that these predispositions are implanted in us by nature and that we are therefore not responsible for their existence.<sup>45</sup> Second, though these predispositions may be toward the good, they are readily twisted into vices: the predisposition to animality (i.e., a merely mechanical self-love, involving drives for self-preservation, reproduction, and sociability) can lead to the *bestial* vices of gluttony, lust, and lawlessness; the predisposition to humanity (i.e., a comparative and potentially competitive self-love, involving *inter alia* a need to gain worth in the eyes of others, up to but not exceeding equal worth, and a need to prevent others from gaining superiority over oneself) can lead to the *cultural* vices of jealousy, rivalry, and even envy and *Schadenfreude*.<sup>46</sup> This is why Allen Wood speaks of the innocent (as opposed to the good) will being in “a natural (if fragile) harmony with the good”: the goodness of our original predispositions to self-love is a delicate one, a primal innocence that is easily corrupted.<sup>47</sup>

The heteronomous and even anomic qualities of prudential reasoning explain why Kant himself neither takes it to exemplify our freedom and rational agency nor makes it a source of our self-worth or a reason for why we should be objects of respect. Rather, it is our *moral* agency that has these qualities, as Kant emphasizes again and again in his writings—for example, when he argues that “a rational being must always be regarded as at the same time *lawgiving*, since otherwise it could not be thought as an *end in itself*.”<sup>48</sup> Kant would almost certainly categorize prudence as a *conditional* good: though it is perhaps conducive to good conduct in most cases, it can be put to immoral uses (e.g., a prudent terrorist), in which case it becomes a bad rather than a good. As Kant says, “the coolness of a scoundrel makes him not only far more dangerous *but also immediately more abominable* in our eyes than we would have taken him to be without.”<sup>49</sup> Respect, whether for ourselves or others, must have an unconditional and permanent basis, one which only our capacity for moral autonomy can provide.

#### *KANTIAN PERSONAL AUTONOMY: A FRESH START*

Given what I have said so far, a *genuinely* Kantian personal autonomy, one that can bridge moral and personal autonomy, would need to meet two apparently inconsistent criteria:

1. it would need to be ultimately motivated not by idiosyncratic and unstable inclinations but rather by pure practical reason, that is, motivated by *objective* reasons, which are uni-

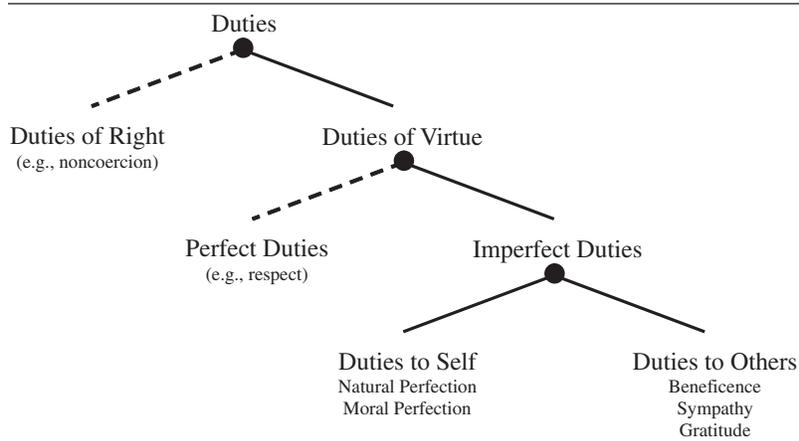
versal in form, bind all finite rational beings, and provide a “unity to our lives” (Schneewind) that is lacking under the mere rule of nature;

2. it would need to make substantial room for *subjective* reasons (i.e., reasons specific to particular agents) in order to make self-creation or self-authorship a real possibility, for autonomy on the Kantian model threatens to reduce “[self-]authorship . . . to a vanishing point as it allows only one set of principles which people can rationally legislate and they are the same for all” (Raz).<sup>50</sup>

Is it even possible for a single conception of personal autonomy to meet both of these criteria, to be not only ultimately objective but also substantially subjective? I shall argue in this section that Kant offers just such a conception in his *Tugendlehre*, or “Doctrine of Virtue”—specifically, in the form of certain imperfect duties of virtue to self (natural perfection) and others (beneficence).

First, however, I want to place these duties in context by very briefly reviewing part of Kant’s taxonomy of duties in his late moral philosophy. (See Figure 1.<sup>51</sup>) Kant divides duties into duties of right and duties of virtue, which are distinguished by (1) the presence of *external enforcement* with respect to actions and omissions in the former and its absence in the latter and by (2) the absence of a requirement to adopt particular *ends* in the former and its presence in the latter.<sup>52</sup> Thus, for example, the duty of noncoercion (proscription of murder, etc.) may be enforced by threat of punishment but does not require those who discharge it to have any particular end (e.g., the right of human beings) in mind when doing so, whereas the duty of respect (proscription of contempt, ridicule, etc.) may not be enforced by threat of punishment but does require those who discharge it to have a particular end (namely, a recognition of the dignity of other human beings) in mind when doing so.<sup>53</sup>

Duties of virtue are further subdivided into perfect and imperfect duties (of virtue).<sup>54</sup> Perfect duties of virtue require particular actions/omissions because these are so closely related to the adoption of particular ends, whereas imperfect duties of virtue allow *latitude* in the choice of actions/omissions so long as the end is still adopted. For example, respect is a perfect duty of virtue, as particular signs of disrespect (e.g., defamation) are to be avoided as simply inconsistent with the end of respecting the dignity of others; beneficence, on the other hand, is an imperfect duty of virtue, because it requires no particular actions/omissions (e.g., giving \$50 to Habitat for Humanity) so long as the end of helping others is still adopted.<sup>55</sup> Finally, the imperfect duties of virtue are further subdivided into (imperfect) duties (of virtue) to self and others. The imperfect duties of virtue include duties to perfect oneself (both naturally and morally) and to love others in a practical manner (through beneficence, sympathy, and gratitude).



**Figure 1: Partial Taxonomy of Duties in Kant's Late Moral Philosophy**

*Natural perfection* of oneself involves the cultivation of one's natural faculties and talents of body, mind, and spirit, both for its own sake and for the sake of myriad possible purposes, moral ones especially.<sup>56</sup> As noted above, the capacity to set ends is what defines humanity; respect for humanity in our own person (specified by the Formula of Humanity as an End-in-Itself) therefore entails the development of this capacity along myriad dimensions.<sup>57</sup> Kant emphasizes, however, that we have great discretion with respect to both *which* talents we should develop and *how much* we should develop them (the extensive and intensive margins, respectively) and that our choice of occupation largely determines which set of skills we develop.<sup>58</sup> *Moral perfection* of oneself requires the cultivation of one's moral disposition, with the ultimate aim of purity, so that respect for the moral law is a sufficient incentive for discharging duty—a purity that may not be attained in this life but only in the next.<sup>59</sup> Kant leaves little room for discretion here, though: he says that this duty is “narrow and perfect with respect to its object” and that one must therefore “strive with all one's might” for moral purity.<sup>60</sup>

Practical love of others is related to willing, not feeling, as feeling cannot be commanded as a duty.<sup>61</sup> What is to be willed here is benevolent *conduct*, as specified in the duties of *beneficence*, *sympathy*, and *gratitude*. Like Kant, I will focus on the first. Beneficence is the duty to promote the happiness of others by making their (permitted) ends our ends as well and by then acting to advance those ends.<sup>62</sup> Kant says “there is still only a negative and not a positive agreement with *humanity as an end in itself* unless everyone also tries, as

far as he can, to further the ends of others.”<sup>63</sup> Respecting the rights of others, in other words, is a necessary but not a sufficient condition for respecting their humanity: if the capacity to set ends defines humanity, then we must help others to achieve their (permissible) ends in order to respect them fully. As was the case with natural perfection, we have a great deal of discretion regarding *how* and *to what extent* we discharge our duty of beneficence. As Kant puts it, “the duty [of beneficence] has in it a latitude for doing more or less, and no specific limits can be assigned to what should be done”; we may, for example, refuse requests for help if we believe they will not promote the happiness of the recipient or if they conflict with our “true needs.”<sup>64</sup>

I can now return to my original claim, namely, that Kant implicitly offers us a conception of personal autonomy in his Doctrine of Virtue in the form of two imperfect duties of virtue: natural perfection of oneself and beneficence toward others. We can see how these two duties constitute a conception of personal autonomy if we consider how they meet the two criteria for a Kantian conception of personal autonomy enumerated above. First, they are both *ultimately* motivated by objective reasons. As we have seen, the duties of self-perfection and the practical love of others follow from a respect for humanity as an end in itself, whether it lies in our own person or in the person of others, and humanity is the highest objective end of pure practical reason. These duties provide architectonic principles for a virtuous life that flow from our own capacity for autonomy and thus help to correct the anomic and heteronomous features of prudential reasoning.

Second, both duties allow subjective reasons to play a substantial role. By leaving so much to discretion, these duties create a great deal of room for self-creation or self-authorship. A virtuous Kantian agent has enormous freedom to fashion his own plan of life—one reflecting his own idiosyncratic “tastes, opinions, ideals, goals, values, and preferences” (Feinberg)—while *at the same time* discharging his duties to perfect himself and advance the happiness of others.<sup>65</sup> For example, self-perfection can be achieved through any number of different occupations, hobbies, and other personal projects that will often be *a* central if not *the* central focus of agents’ plans of life. More often than not, advancing the happiness of others plays a similarly large role, whether it is achieved in intimate settings (e.g., through one’s support for family, friends, and colleagues) or in more impersonal, institutionalized ways (e.g., gifts of cash or labor to organized charities). By choosing among these different means to (ultimate) objective ends, agents can fashion unique and satisfying lives for themselves, authentic lives of their own making but also lives having a moral point. In other words, the shape of such lives can be responsive to *subjective reasons* (e.g., choosing one occupation over another because of one’s tastes and/or natural abilities) without ceasing to be about

*objective ends*; the pursuit of such ends validates or redeems supporting subjective reasons, so to speak, despite their heteronomous nature. As Allen Wood says, “if I am a decent person, I will choose to give my life meaning by pursuing some set of ends that fall under the general description ‘my own perfection’ and ‘the happiness of others.’ Where that is so, morality underwrites our ground projects, regarding them as morally meritorious.”<sup>66</sup>

None of this is meant to deny, of course, that Kantian agents may adopt *merely* subjective ends that are consistent with both right and the perfect duties of virtue. An example might be a certain type of consumption activity: although some kinds of consumption have a self-improving quality to them (e.g., reading challenging novels), others will lack this quality and thus qualify as merely subjective ends (e.g., reading the *Beetle Bailey* comic strip).<sup>67</sup> Kantian duties of virtue are not “maximalist,” demanding that moral agents sacrifice all their merely subjective ends in order to perfect themselves and aid others. These concerns must play *some* role in the life of a virtuous agent, of course, but he can limit them without blame for the sake of merely subjective ends, including especially his “true needs” (for food, clothing, shelter, companionship, etc.).<sup>68</sup>

The conception of personal autonomy implicit in Kant’s Doctrine of Virtue avoids a problem that is endemic to the personal-autonomy literature: autonomy as it is conceived in this literature is consistent with unethical, even criminal behavior, which leads one to question its independent value. As I noted at the end of the last section, Kant considers prudence—the first of the Kantian conceptions of personal autonomy we examined—to be a merely *conditional* good: as Kant says, “the coolness of a scoundrel makes him not only far more dangerous *but also immediately more abominable* in our eyes than we would have taken him to be without.”<sup>69</sup> The same point would seem to apply to personal autonomy more generally, but the Kantian personal autonomy that we explored in this section is immune to such criticisms. Far from condoning immoral behavior, this version is itself a moral enterprise, directing agents to advance the objective end of humanity in themselves and others.

Such a solution appears to create a new problem, however. Autonomy has typically meant a certain independence from constraint, so how then can a meaningful conception of autonomy be constituted by the constraints of moral duty? The answer to this question is that these constraints are themselves a product of self-legislation; autonomy is not inconsistent with constraints *per se* but rather with constraints imposed on an agent by someone else. In the case of Kantian personal autonomy, this self-legislation has two facets. First, the imperfect duties of self-perfection and beneficence are creations of our own pure practical reason: maxims inconsistent with these two

duties cannot be universalized (as they lead to a contradiction in will) and therefore violate the first formulation of the categorical imperative, which itself follows from our autonomous will.<sup>70</sup> Second, despite our discretion with respect to how and to what degree we carry out these duties, we must create plans of life that are at least minimally responsive to them, plans that are more idiosyncratic and particularistic forms of self-legislation. This dual self-legislation combines elements of moral and personal autonomy, respectively, into a unified whole.

The genuinely Kantian personal autonomy just described indicates that the sharp distinction between Kantian moral autonomy and contemporary personal autonomy—a distinction made, as we have seen, by contemporary Kantians and personal-autonomy theorists alike—is overdrawn. More generally, the argument of this section has revealed a potential connection between what had previously been regarded as fundamentally different conceptions of autonomy, suggesting that there may be other connections as well and offering us the prospect of a unified theory of autonomy. Furthermore, our search for conceptions of personal autonomy in Kant's moral philosophy has revealed in his work an unnoticed (or perhaps undernoticed) concern for self-authorship, Raz's claims to the contrary notwithstanding.<sup>71</sup> What may have seemed a hopelessly anachronistic enterprise at the start has proven both philosophically and exegetically useful.

*CONTEMPORARY PHILOSOPHICAL  
APPLICATION: JOHN RAWLS*

In addition to improving our understanding of Kant's own moral theory and the relationship among the various conceptions of autonomy, Kantian personal autonomy provides an additional, much-needed line of defense against criticisms of *contemporary* Kantian theories, especially that of John Rawls. Two criticisms of Rawls's theory are of particular interest here: the first, which I shall call the "Discontinuity Criticism," comes from inside the liberal camp; the second, which I shall call the "Arbitrariness Criticism," comes from outside it. I shall describe each criticism in turn, showing how Kantian personal autonomy provides Rawls with a powerful response to his critics—though probably not one that the later Rawls could have employed, as we shall see.

Jeremy Waldron states the Discontinuity Criticism succinctly when he says that Rawls "has to explain why individuals are willing to subordinate their pursuit of the good to principles of right, and how that subordination is possible given the discontinuity between moral autonomy and the pursuit of

an individual conception of the good,” a discontinuity that threatens to make strictures of right seem “alien . . . from the personal point of view.”<sup>72</sup> Consider the religious fundamentalist, for example, who believes those outside the church will suffer eternal hellfire and that physical force is therefore justified to bring them into the fold.<sup>73</sup> He may come to accept the arguments for liberty of conscience, but perhaps only as an alien imposition bearing little relation to his deeply held religious faith. This example presents the discontinuity in its starkest form.

Rawls, of course, is perfectly aware of the possibility of such discontinuity: it motivates his complex and at times obscure “congruence argument,” which tries to show that “the disposition to take up and be guided by the standpoint of justice accords with the individual’s good.”<sup>74</sup> This argument knits together various claims defended throughout *Theory*, including the propositions that a well-ordered society made possible by justice greatly contributes to our pursuit of the good and that acting justly expresses our most fundamental desire *qua* free and equal rational beings. Waldron believes that the congruence argument is unsuccessful, a position that is taken by many scholars and apparently by Rawls himself.<sup>75</sup> Even those scholars who defend the argument or some reconstruction of it (e.g., Samuel Freeman) recognize its limitations and many critics.<sup>76</sup>

Without taking a position on the merits of the congruence argument (whether as originally presented or as reconstructed by Freeman *inter alia*), we can readily see how Kantian personal autonomy offers a different and perhaps more compelling way to bridge the gap between justice and goodness. Respect for humanity as an end in itself requires us to discharge both duties of right and duties of virtue, and as we have seen, certain imperfect duties of virtue to self and others constitute Kantian personal autonomy. Therefore, when we incorporate objective ends (specifically, self-perfection and the happiness of others) into our conceptions of the good and the plans of life that realize them—as virtue requires us to do—we demonstrate the very same respect for humanity that we show by fulfilling our duties of justice. In short, there need be no sharp discontinuity between justice and goodness, between moral and personal autonomy, if goodness and personal autonomy are interpreted in a Kantian fashion. Of course, such a move may require us to adopt a critical attitude toward certain existing conceptions of the good (e.g., religious fundamentalism), as Kantian morality is consistent with many different religious and ethical traditions but is not and indeed could not be neutral toward all conceptions of the good.<sup>77</sup> As we shall soon see, this feature of Kantian personal autonomy makes it unlikely that Rawls could have accepted it following his so-called “political turn.”

Unlike the Discontinuity Criticism, the Arbitrariness Criticism comes not from within the liberal camp but rather from communitarian adversaries of Rawls, most notably Michael Sandel. Sandel famously criticizes Rawls for his conception of the moral agent. He argues, among other things, that the “unencumbered and essentially dispossessed” self of Rawlsian theory has no basis for choice in matters of goodness (as opposed to justice): if the “self is prior to the ends which are affirmed by it,” as Rawls says, then “deliberation about ends can only be an exercise in arbitrariness,” as there are no *preexisting* materials (e.g., preferences, values, goals) upon which to base such deliberation.<sup>78</sup> In other words, the priority of the self over its ends implies that prudential reasoning must involve groundless choice or a kind of deliberative bootstrapping.

The Arbitrariness Criticism has itself been roundly criticized, most notably by Will Kymlicka, who responds to Sandel by noting that the priority of the self asserted by Rawls means only that our given ends can be criticized and revised, *not* that we can practice groundless choice or ever envision ourselves as fully “unencumbered.”<sup>79</sup> Though some of Sandel’s claims here are indeed quite questionable, the arbitrariness charge in particular is harder to rebut because Rawls himself acknowledges this attribute of his theory of deliberative rationality: he speaks of the “arbitrary features of plans of life” as well as the “indeterminacy of the conception of the good.”<sup>80</sup> Rawls considers such indeterminacy to be “relatively innocuous” because, given the priority of the right over the good, the potential arbitrariness of the good does not effect the secure status of “men’s claims on one another.”<sup>81</sup> Sandel’s concern regarding this indeterminacy seems slightly different, however: if men’s choices are bounded *only* by the (admittedly secure) principles of justice and rational choice, the field of choice will remain dizzying large, leaving men who live in a fluid, plural society susceptible to the disorientation and normlessness of Durkheimian *anomie*.

Rawls might respond to this concern with an intellectual-division-of-labor argument—to wit, that securing rights is the business of justice and providing purpose is the business of the various moral and religious doctrines of the good present in any diverse society—but such a response would merely strengthen the force of the Discontinuity Criticism. Kantian personal autonomy provides Rawls with a way of confronting Sandel’s concerns directly. Rather than simply adding new constraints to those already provided by principles of justice and rational choice, Kantian personal autonomy offers a set of objective ends to help agents orient themselves both personally and morally when they construct their plans of life. Indeterminacy, of course, remains an issue (e.g., one must still decide how and to what extent to perfect oneself

and advance the happiness of others), but these objective ends, if adopted by the conscientious agent, provide at least *some* positive guidance to augment the mostly negative, prohibitive rules of right and prudence.<sup>82</sup> In short, Rawls's best response to the Arbitrariness Criticism might be a *more thoroughly* Kantian liberalism rather than the retreat from Kantianism into communitarianism suggested by Sandel.

Such a response would have been difficult for Rawls, however, after his so-called "political turn": he says in his late writings that the "full autonomy of *political* life must be distinguished from the *ethical* values of autonomy and individuality, which may apply to the whole of life, both social and individual, as expressed by the comprehensive liberalisms of Kant and Mill"; thus, it is highly unlikely that Rawls could have endorsed Kantian personal autonomy at this stage given its inextricable connection to Kant's comprehensive liberalism.<sup>83</sup> Rawls's "political turn" was motivated by his concerns regarding the public justification and stability of justice as fairness: given the fact of reasonable pluralism, liberal societies will assuredly be populated by adherents of non-Kantian reasonable comprehensive doctrines who will be unable to endorse a justification of political principles grounded on a Kantian understanding of autonomy and whose commitment to those principles may therefore prove unstable. His proposed solution to this problem is to offer a justification of these principles that does not rely on the idiosyncratic assumptions of any given reasonable comprehensive doctrine but rather on the "fundamental ideas in a democratic political culture" that are shared by *all* such doctrines.<sup>84</sup> The later Rawls would therefore have regarded Kantian personal autonomy as incompatible with *political* liberalism, as its strong assumptions about autonomy, virtue, and objective ends cannot be publicly justified, that is, cannot be part of an overlapping consensus of the diverse reasonable comprehensive doctrines in a liberal society.

This evolution of Rawls's theory apparently renders any discussion of his potential adoption of Kantian personal autonomy moot: why should we care whether Kantian personal autonomy would provide an additional line of defense for a superseded theory? I say "apparently," because political liberalism itself has been subjected to searching criticism, and if Rawls's "political turn" turns out to be mistaken, then an interest in the structure of his earlier, comprehensive liberalism might be justified.<sup>85</sup> One concern about political liberalism has been raised by John Tomasi, who argues that "political liberals cannot hope to adopt a wholly new motivational base for their view and yet have the content of justice remain exactly the same as before. . . ."<sup>86</sup> More specifically, the weaker assumptions of political liberalism, which allow a diversity of reasonable comprehensive doctrines to participate in an overlapping

consensus, may simply yield *weaker results* in terms of the political principles that can be supported.

In a recent article, for example, I argue that the lexical priority of liberty can be defended only with a robustly Kantian conception of the person, one that would be difficult for other reasonable comprehensive doctrines to accept.<sup>87</sup> Briefly, Rawls's best defense of the priority of liberty takes our highest-order interest to be the preservation of our rationality (understood as our capacity to form and revise a conception of the good and an associated plan of life) and the preconditions of its exercise, especially the basic liberties. As a facet of our autonomy, rationality takes absolute priority over our other interests (such as material consumption), which is why the basic liberties that support it take absolute priority over other social primary goods. The strong conception of the person underlying these claims appears inconsistent with certain reasonable comprehensive doctrines, which is presumably why Rawls explicitly excludes the classical utilitarianism of Bentham and Sidgwick from his earliest versions of the overlapping consensus.<sup>88</sup>

Rawls's defense of the difference principle in *Theory* relies upon similarly strong assumptions about the structure of conceptions of the good, assumptions that may be in tension with many or most reasonable comprehensive doctrines. In Section 26 of *Theory*, for example, Rawls points out that one of the three conditions for the adoption of maximin is that "the person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule"; he adds shortly thereafter that the difference principle "guarantees a satisfactory minimum. There may be, on reflection, little reason for trying to do better."<sup>89</sup> It is unclear why liberal utilitarians and other partisans of non-Kantian but reasonable comprehensive doctrines would endorse this kind of asceticism, though it *would* be consistent with (if not endorsed by) a Kantian doctrine. If the priority of liberty and the difference principle are indeed what give justice as fairness its "force," as Rawls asserts in *Theory*, then the move from a comprehensive to a political liberalism might seriously undermine the defense of its two most distinctive elements.<sup>90</sup>

Thus, in order to argue more effectively for justice as fairness, Rawls (or rather his followers) might need to return to *Theory*'s comprehensive Kantian liberalism. *If* such a return is seriously entertained, then the question of whether the earlier Rawls could have adopted Kantian personal autonomy as a way of deflecting Waldron and Sandel's criticisms becomes more germane. He does in *Theory* leave open the possibility at least of constructing a Rawlsian Doctrine of Virtue:

Justice as fairness is not a complete contract theory. For it is clear that the contractarian idea can be extended to the choice of more or less an entire ethical system, that is, to a system including principles for all the virtues and not only for justice. Now for the most part I shall consider only principles of justice and others closely related to them; I make no attempt to discuss the virtues in a systematic way. Obviously if justice as fairness succeeds reasonably well, a next step would be to study the more general view suggested by the name "rightness as fairness."<sup>91</sup>

Rawls later goes on to sketch the sort of principles that might apply to individuals, especially with regard to natural duties (e.g., mutual aid and mutual respect) and supererogatory actions (e.g., beneficence), but his discussion there is admittedly and intentionally incomplete.<sup>92</sup>

Were Rawls actually to have carried out such an extension of his theory, he would likely have found much in Kant's Doctrine of Virtue congenial. Both base their theories on a conception of man as a "free and equal rational being," and both seek to discover principles that "best manifest this freedom in [human] community [and that] most fully reveal [our] independence from natural contingencies and social accident."<sup>93</sup> A Rawlsian Doctrine of Virtue grounded and motivated in this fashion would surely have placed an emphasis, as Kant's did, on the systematic development of the human capacity to set ends, not only because the exercise of this capacity is so likely to be hampered (in ourselves as well as others) by the "natural contingencies and social accidents" of which Rawls spoke but also because its development features so strongly in his theory of justice. Such an extension of Rawls's theory to a complete ethical system would have put him in a better position to respond to the kinds of criticisms leveled by Waldron and Sandel, but as we know he chose to develop his theory in another direction entirely.

If Rawls's theory *were* supplemented by something like Kantian personal autonomy, would it then qualify as a variety of liberal perfectionism like the theories of William Galston and Joseph Raz?<sup>94</sup> On Rawls's *own* understanding of perfectionism, it would not. Rawls sees perfectionism as a teleological theory that maximizes the good, understood in this context as "the realization of human excellence in the various forms of culture."<sup>95</sup> He explicitly says that "justice as fairness is not a maximizing theory" and that the human excellences constituting moral personality (i.e., the capacities for a conception of the good and for a sense of justice) need only meet a low "essential minimum . . . ordinarily realized in due course."<sup>96</sup> Rawls's theory would not be maximizing even if it were supplemented by duties of self-perfection like those found in Kantian personal autonomy: as we have seen, natural perfection of oneself is an imperfect duty, discretionary with respect to both which talents are developed and *how much* they are developed.

Rawls's definition of perfectionism is arguably too narrow, however. One indication of this is that the theories of self-identifying liberal perfectionists such as Galston and Raz would probably fail to be perfectionist on Rawls's account: the radical value pluralism of their theories and the absence of clear weighting or ranking rules make it difficult to categorize them as maximizing.<sup>97</sup> If we understand perfectionism in a somewhat looser sense, however, as concerning itself with the promotion if not the maximization of certain human excellences, then his theory is plausibly thought of as perfectionist, especially if it is supplemented by something like Kantian personal autonomy. As noted above, Rawls's original theory recognizes the need to develop (if only to a modest degree) the human capacities for rationality and reasonableness, and its commitment to the lexical priority of fair equality of opportunity arguably reveals a moderate perfectionism, as the function of such priority is to privilege the pursuit of virtuosity in the various positions and offices of the basic structure.<sup>98</sup> If a doctrine of virtue that included imperfect duties of (natural) self-perfection were added to Rawls's theory, its perfectionism would be even more striking, at least by the looser definition of the term.

### CONCLUSION

I began the central part of my paper by asking whether Kantian prudence might qualify as a conception of personal autonomy. I went on to discover a series of striking similarities between the two—including a hierarchical integration of the inclinations, a combination of idiosyncratic self-authorship with deliberative rationality, and a dependence on such ancillary virtues as self-control—but finally concluded that it qualifies as a Kantian personal autonomy only in the first, weaker sense of that term. It failed to qualify in the second, deeper sense because Kant does not and in fact cannot view prudence as a species of autonomy: the disorderliness and variability of our desires and their exclusive origin in nature and in corrupt society imply that the prudential reason that attempts to govern them will at best be heteronomous and at worst anomic. I then began anew by considering whether the elements of a conception of personal autonomy could be found in Kant's "Doctrine of Virtue"—specifically, in the imperfect duties of self-perfection and the practical love of others. These duties indeed met the two criteria I specified for a Kantian personal autonomy because they are responsive to the subjective reasons of individual agents while still being directed toward the objective end of humanity. I concluded by showing that Kantian personal autonomy, if adopted by Rawls, could provide an additional line of defense against two

criticisms of his theory of autonomous agency by Waldron and Sandel. By giving Rawls a means to integrate personal and moral autonomy and to provide positive guidance to conscientious moral agents, Kantian personal autonomy would in fact allow him to deflect these criticisms, but only at the cost of developing his own Doctrine of Virtue, something that would be difficult after the “political turn” but quite attractive if a return to the comprehensive Kantian liberalism of *Theory* were entertained.

Over the course of the essay I have emphasized the philosophical and exegetical value of this comparative exercise, but it is valuable in other ways as well. As James Schmidt has observed, critics of the Enlightenment commonly condemn it for “reducing society to a series of contracts between otherwise isolated individuals,” a criticism also applied to contemporary liberalism.<sup>99</sup> As applied to Kant and many other Enlightenment figures (e.g., David Hume and Adam Smith), the criticism is absurd. Some versions of contemporary liberalism, on the other hand, have made themselves vulnerable to such a charge by either ignoring virtue or by focusing exclusively on its procedural political dimension (e.g., toleration, civility, reasonableness, etc.).<sup>100</sup> Kantian personal autonomy offers one way (though certainly not the only way) for practitioners of these versions of liberalism to reengage with matters of virtue; it offers personal autonomy with a *moral point*, without abandoning liberal principles. The cure to what ails contemporary liberalism, in short, is not a disengagement from our Enlightenment heritage but rather a more systematic engagement with it.

### NOTES

1. John Stuart Mill, *On Liberty and Other Essays*, ed. John Gray (New York: Oxford University Press, 1991), 63.

2. Jeremy Waldron, “Moral Autonomy and Personal Autonomy,” in *Autonomy and the Challenges to Liberalism*, eds. John Christman and Joel Anderson (Cambridge: Cambridge University Press, 2005), 308–14.

3. Joseph Raz, *The Morality of Freedom* (New York: Oxford University Press, 1986), 370n. Cf. Joel Feinberg, “Autonomy,” in *The Inner Citadel: Essays on Individual Autonomy*, ed. John Christman (New York: Oxford University Press, 1989), 35–6, 44.

4. Onora O’Neill, *Constructions of Reason: Explorations of Kant’s Practical Philosophy* (Cambridge: Cambridge University Press, 1989), 66, 75. Cf. Thomas Hill, “The Kantian Conception of Autonomy,” in *The Inner Citadel*, 92–5.

5. Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988), 9. On the concept/conception distinction, see H. L. A. Hart, *The Concept of Law* (Oxford: The Clarendon Press, 1961), 155–9, and John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1999 [Revised Edition]), 5, 8–9.

6. In a similar spirit, Waldron says that “what is called for is some sort of moderate or intermediate position. This chapter has not sought to identify that third way, but I hope it has helped

illuminate some of the perils as well as some of the advantages of a distinction between personal and moral autonomy." Waldron, "Moral Autonomy and Personal Autonomy," 325.

7. Raz, *The Morality of Freedom*, 369.

8. *Ibid.*, 371, 375–6, 382.

9. John Rawls, "Kantian Constructivism in Moral Theory," in *Collected Papers*, ed. Samuel Freeman (Cambridge, MA: Harvard University Press, 1999), 308, 319.

10. *Ibid.*, 316. Cf. Rawls, *A Theory of Justice*, §63–4.

11. Dworkin, *The Theory and Practice of Autonomy*, 20, and Gerald Dworkin, "Autonomy," in *A Companion to Contemporary Political Philosophy*, eds. Robert E. Goodin and Philip Pettit (Oxford: Blackwell, 1993). Cf. Harry Frankfurt, "Freedom of the Will and the Concept of the Person," in Harry Frankfurt, *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press, 1988), 15, 19. Also see Michael Bratman, "Autonomy and Hierarchy," *Social Philosophy and Policy* 20 (2003): 156–76.

12. Rawls, *A Theory of Justice*, 360–1.

13. John Christman, "Autonomy in Moral and Political Philosophy," in *Stanford Encyclopedia of Philosophy (Fall 2003 Edition)*, ed. Edward N. Zalta, <http://plato.stanford.edu/archives/fall2003/entries/autonomy-moral/>. Also see John Christman, "Constructing the Inner Citadel: Recent Work on the Concept of Autonomy," *Ethics* 99 (1988): 109–24.

14. Feinberg, "Autonomy," 30–43.

15. Alfred Mele, *Autonomous Agents: From Self-Control to Autonomy* (Oxford: Oxford University Press, 2001).

16. Dworkin, *The Theory and Practice of Autonomy*, 6, 9.

17. David Johnston, *The Idea of a Liberal Theory: A Critique and Reconstruction* (Princeton: Princeton University Press, 1994), 76–7; Raz, *The Morality of Freedom*, 380.

18. Waldron, "Moral Autonomy and Personal Autonomy," 311, 314.

19. Immanuel Kant, *Practical Philosophy*, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1996), 54 [GMM 4:399], 69n [GMM 4:414]. *N.B.*: Throughout the endnotes I will supplement the page references to this edition of Kant with (i) abbreviations for the works from which they were drawn plus (ii) the relevant volume and page references to the standard German edition of Kant's works, *Kant's Gesammelte Schriften*, ed. Royal Prussian Academy of Sciences (Berlin: Georg Reimer, 1900). Here is a complete list of the abbreviations I use: GMM=*Groundwork of the Metaphysics of Morals*; MM=*Metaphysics of Morals*; CPrR=*Critique of Practical Reason*; T&P="On the common saying: That may be correct in theory, but it is of no use in practice"; Rel=*Religion within the Boundaries of Mere Reason*.

20. *Ibid.*, 199 [CPrR 5:73]; Immanuel Kant, *Religion within the Boundaries of Mere Reason*, eds. Allen Wood and George di Giovanni (Cambridge: Cambridge University Press, 1998), 78 [Rel 6:58].

21. See H. J. Paton, *The Categorical Imperative: A Study in Kant's Moral Philosophy* (Philadelphia: University of Pennsylvania Press, 1947), 85–7.

22. Rawls, *A Theory of Justice*, pp. 360–1. Cf. Plato, *Republic*, trans. Allan Bloom (New York: Basic Books, 1991), 272 [589a–c].

23. See Mill, *On Liberty and Other Essays*, 66: "Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

24. Kant, *Practical Philosophy*, 54 [GMM 4:399].

25. *Ibid.*, 157–8 [CPrR 5:24]; cf. *ibid.*, 371 [MM 6:216].

26. *Ibid.*, 71 [GMM 4:419].

27. *Ibid.*, 371 [MM 6:216]. Also see Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 111–2.
28. Kant, *Practical Philosophy*, 562–3 [MM 6:441–2].
29. *Ibid.*, 49–50 [GMM 4:394]. Cf. Kant, *Religion*, 51n [Rel 6:27].
30. Paul Guyer, *Kant on Freedom, Law, and Happiness* (Cambridge: Cambridge University Press, 2000), 406. Cf. Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 164–5.
31. Kant, *Practical Philosophy*, 228–9 [CPrR 5:111].
32. Kant, *Religion*, 50–1 [Rel 6:27], 78 [Rel 6:58]. Cf. Kant, *Practical Philosophy*, 79 [GMM 4:428], for a different perspective. These predispositions are to mechanical self-love (animality) and comparative self-love (humanity), which are roughly equivalent to Rousseau's concepts of *amour de soi-même* and *amour propre*, respectively. See Jean-Jacques Rousseau, *The Discourses and Other Early Political Writings*, ed. Victor Gourevitch (Cambridge: Cambridge University Press, 1997), 218 [Second Discourse, Rousseau's Note 15].
33. Dworkin, *The Theory and Practice of Autonomy*, 31.
34. Allison, *Kant's Theory of Freedom*, 88–9; Kant, *Practical Philosophy*, 70 [GMM 4:417].
35. Korsgaard, *Creating the Kingdom of Ends*, Chapter 4; Kant, *Practical Philosophy*, 80 [GMM 4:429], 522 [MM 6:392]. Cf. Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), 120: "Preserving and respecting rational nature means preserving and respecting it in all its functions, not merely in its moral function of giving and obeying moral laws. Furthering rational nature requires furthering all the (morally permissible) ends it sets, not merely the ends it sets in response to duty."
36. *Ibid.*, 74.
37. Kant, *Practical Philosophy*, 70–1 [GMM 4:418–9].
38. *Ibid.*, 54 [GMM 4:399], 71 [GMM 4:419].
39. J. B. Schneewind, "Kant and Stoic Ethics," in *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, eds. Stephen Engstrom and Jennifer Whiting (Cambridge: Cambridge University Press, 1996), 290.
40. Kant, *Practical Philosophy*, 159–60 [CPrR 5:26].
41. *Ibid.*, 285n [T&P 8:284] (emphasis added).
42. *Ibid.*, 235 [CPrR 5:118], 268 [CPrR 5:161]. Cf. *ibid.*, 79 [GMM 4:428].
43. *Ibid.*, 189–90 [CPrR 5:62].
44. Rawls nonetheless defends Kant against this charge, arguing that his moral psychology is less Manichaeian than Augustinian. See John Rawls, *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University Press, 2000), 303–6.
45. Kant, *Religion*, 57–8 [Rel 6:35].
46. *Ibid.*, 51 [Rel 6:27]. Also see endnote on Rousseau above.
47. Wood, *Kant's Ethical Thought*, 47. If Kant's belief in the natural and/or social genesis of all inclinations appears implausible, see Andrew Reath's interpretation of Kant on the nature of heteronomy and the genealogy of desire in "Hedonism, Heteronomy and Kant's Principle of Happiness," *Pacific Philosophical Quarterly* 70 (1989): 42–72, especially 48, 60.
48. Kant, *Practical Philosophy*, 84 [GMM 4:435].
49. *Ibid.*, 50 [GMM 4:394] (emphasis added). Many if not most theorists of personal autonomy recognize the conditionality of its goodness. See, for example, Raz, *The Morality of Freedom*, 380–1; Frankfurt, *The Importance of What We Care About*, 19n; Dworkin, *The Theory and Practice of Autonomy*, 29; and Feinberg, "Autonomy," 39, 43.
50. Schneewind, "Kant and Stoic Ethics," 290; Raz, *The Morality of Freedom*, 370n.
51. Based on diagrams in Kant, *Practical Philosophy*, 395 [MM 6:240], 527 [MM 6:398].

52. *Ibid.*, 512–5 [MM 6:380–3]. One can also discharge one’s duties of right virtuously (i.e., meritoriously) if one makes “the right of humanity, or also the right of human beings, one’s *end*.” *Ibid.*, 521–2 [MM 6:391].

53. *Ibid.*, 474–6 [MM 6:333–6], 579–83 [MM 6:462–8].

54. On the distinction between perfect and imperfect duties in general, see *ibid.*, 521–2 [MM 6:390–2].

55. *Ibid.*, 571–3 [MM 6:452–5]. Though the duty of beneficence requires no particular actions, a *complete absence* of beneficent actions would be a violation of the duty. See *ibid.*, 75 [GMM 4:424], 81 [GMM 4:431], 572 [MM 6:453]. Also, all duties of right are perfect duties: one cannot impose an end by use of coercion, though one can use coercion to guarantee the performance or omission of an action. See *ibid.*, 521 [MM 6:390–1].

56. *Ibid.*, 74–5 [GMM 4:423], 518 [MM 6:387], 522–3 [MM 6:392], 565–6 [MM 6:445–6]. Kant explicitly rules out the perfection of *others* as a duty: he says that perfection is “something that only the other himself can do,” for unless the other himself sets it as an end, the cultivation of his talents will fail to be meritorious. Again, one can force actions but not ends, and only the free adoption of virtuous ends can earn merit. *Ibid.*, 517–8 [MM 6:386].

57. *Ibid.*, 80 [GMM 4:429], 522–3 [MM 6:392]. For a different approach to justifying natural perfection as a duty, see O’Neill, *Constructions of Reason*, 98–101.

58. Kant, *Practical Philosophy*, 523 [MM 6:392], 566 [MM 6:446]. On the precise nature of the relationship between natural perfection and occupation, see Robert S. Taylor, “Self-Realization and the Priority of Fair Equality of Opportunity,” *Journal of Moral Philosophy* 1 (2004): 333–47, here 342–3.

59. Kant, *Practical Philosophy*, 238–9 [CPrR 5:122–4], 518–9 [MM 6:387], 523 [MM 6:393], 566–7 [MM 6:447].

60. *Ibid.*, 523 [MM 6:393], 567 [6:447].

61. *Ibid.*, 530 [MM 6:402], 569 [MM 6:450]. Kant rejects the idea of a duty of practical love toward *oneself*: the promotion of one’s own happiness is at most an *indirect* duty and only insofar as its promotion is necessary to ward off poverty, pain, and so on, as “great temptations to violate one’s duty.” That is, the promotion of one’s own happiness is, from a moral point of view, only useful as a means to promote (adherence to) morality. *Ibid.*, 517–8 [MM 6:386].

62. *Ibid.*, 519 [MM 6:388], 524 [MM 6:394], 571–3 [MM 6:453–5].

63. *Ibid.*, 81 [GMM 4:431]. Also see Barbara Herman, *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993), Chapter 3, as well as O’Neill, *Constructions of Reason*, 98–101.

64. *Ibid.*, 519 [MM 6:388], 524 [MM 6:394].

65. Feinberg, “Autonomy,” 32.

66. Wood, *Kant’s Ethical Thought*, 328–9. Cf. Reath, “Hedonism, Heteronomy and Kant’s Principle of Happiness,” 66–8.

67. On consumption as a self-actualizing (but not self-realizing) activity, see Jon Elster, “Self-Realization in Work and Politics: The Marxist Conception of the Good Life,” *Social Philosophy & Policy* 3 (1986): 103, 106.

68. Kant, *Practical Philosophy*, 524 [MM 6:394], 555–6 [MM 6:432–3], 572 [MM 6:453]. For a discussion of “true needs,” see Rawls, *Lectures on the History of Moral Philosophy*, 173–6, 221, 232–4. On the larger issue of setting ends in a virtuous life, see Wood, *Kant’s Ethical Thought*, 325.

69. Kant, *Practical Philosophy*, 50 [GMM 4:394] (emphasis added). Cf. Christman, “Constructing the Inner Citadel,” 120.

70. Kant, *Practical Philosophy*, 74–5 [GMM 4:423–4], 94–5 [GMM 4:446–7].

71. Raz, *The Morality of Freedom*, 370n.

72. Waldron, "Moral Autonomy and Personal Autonomy," 308, 319.

73. Cf. Luke 14:23 ("Go out into the highways and hedges, and compel them to come in, that my house may be filled.") and St. Augustine's use of it to defend the persecution of heretics in St. Augustine, *The Political Writings*, ed. Henry Paolucci (Washington, DC: Regnery Publishing, 1996), 193.

74. Rawls, *A Theory of Justice*, 497. Rawls summarizes the myriad strands of the congruence argument in §86 ("The Good of the Sense of Justice").

75. Brian Barry says of the congruence argument that "the only thing to do with it is to follow the course followed virtually unanimously by commentators on *A Theory of Justice* and forget about it." See Brian Barry, "John Rawls and the Search for Stability," *Ethics* 105 (1995): 915n. Rawls's doubts are expressed in *Political Liberalism* (New York: Columbia University Press, 1993), xv–xvii, though the focus here is on the larger argument in Part III of *A Theory of Justice* about the stability of a well-ordered society rather than on the congruence argument proper, which is merely one component (albeit a central one) of the larger argument. Concerns about the stability of justice as fairness were the primary motivation for Rawls's so-called "political turn," as we shall soon see.

76. Freeman attempts to reconstruct Rawls's congruence argument in "Congruence and the Good of Justice," in *The Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge: Cambridge University Press, 2003), 277–315. On the argument's critics and limitations, see especially 277–8, 292, and 303–8.

77. For example, Kant provides "the supreme principle of all scriptural exegesis" in *Religion*, which requires us to interpret scriptural passages so that they are consistent with our obligations of justice and virtue (*Religion*, 119–20 [Rel 6:112]). Consequently, Kant would reject any religious doctrine that interpreted scripture (e.g., Luke 14:23) in such a way as to endorse the persecution of heretics or unbelievers, as such persecution is inconsistent with our prior political duty of religious toleration.

78. Rawls, *A Theory of Justice*, 491; Michael Sandel, *Liberalism and the Limits of Justice* (Cambridge, MA: Harvard University Press, 1982), 180.

79. Will Kymlicka, *Contemporary Political Philosophy: An Introduction* (Oxford: Oxford University Press, 2002 [Second Edition]), 224–8.

80. Rawls, *A Theory of Justice*, 395, 494.

81. *Ibid.*, 494.

82. I may be overstating the residual indeterminacy here, as principles of deliberative rationality may interact with objective ends in such a way as to constrain the choice set further. For example, one principle of deliberative rationality is surely "inventorying," that is, systematically examining our various capacities, preferences, and so on, before we make (important) decisions. If we do this before engaging in self-perfection, we may be able to rule out certain plans of self-development due to the lack of relevant capacities—singing lessons, for instance, are probably not the best path to self-improvement for the congenitally tone deaf. (Special thanks to Brad Inwood for pointing this out.)

83. Rawls, *Political Liberalism*, 78 (emphasis added). For a brief discussion of the key differences between *A Theory of Justice* and *Political Liberalism*, see John Rawls, *The Law of Peoples with "The Idea of Public Reason Revisited"* (Cambridge, MA: Harvard University Press, 1999), 179–80; also see Samuel Freeman's introduction to *The Cambridge Companion to Rawls*, 28–44.

84. Rawls, *Political Liberalism*, 167.

85. Brian Barry, for example, has been a persistent critic of Rawls's "political turn." To quote Barry on this topic: "I believe that, as time goes on, *A Theory of Justice* will stand out with increasing clarity as by far the most significant contribution to political philosophy produced in

this century. Only one thing threatens to obscure that achievement: the publication of *Political Liberalism*." Barry, "John Rawls and the Search for Stability," 915.

86. John Tomasi, *Liberalism Beyond Justice: Citizens, Society, and the Boundaries of Political Theory* (Princeton, NJ: Princeton University Press, 2001), 127.

87. Robert S. Taylor, "Rawls's Defense of the Priority of Liberty: A Kantian Reconstruction," *Philosophy & Public Affairs* 31 (2003): 246–71.

88. Rawls, "The Idea of an Overlapping Consensus," in *Collected Papers*, 433–4. Rawls later changes his mind about this exclusion (*Political Liberalism*, 170), though the reason for this change appears to be his conflation of constitutional and overlapping consensus; see Taylor, "Rawls's Defense of the Priority of Liberty," 268–70.

89. Rawls, *A Theory of Justice*, 134–5.

90. *Ibid.*, 220.

91. *Ibid.*, 15.

92. *Ibid.*, 94, 98–101. See §18–9 in general.

93. *Ibid.*, 225.

94. William Galston, *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State* (Cambridge: Cambridge University Press, 1991); Raz, *The Morality of Freedom*.

95. Rawls, *A Theory of Justice*, 21–2.

96. *Ibid.*, 442, 445.

97. Galston, *Liberal Purposes*, Chapter 8, and Raz, *The Morality of Freedom*, Chapters 13 and 14.

98. Taylor, "Self-Realization and the Priority of Fair Equality of Opportunity."

99. James Schmidt, "What Enlightenment Project?" *Political Theory* 28 (2000): 735.

100. For example, see Rawls, *Political Liberalism*, 194, 199.

*Robert S. Taylor is Assistant Professor of Political Science at the University of California–Davis. He has recently published on the priorities of liberty and fair equality of opportunity in Rawls's "justice as fairness" (Philosophy & Public Affairs and Journal of Moral Philosophy) and on self-ownership (Journal of Political Philosophy and Social Theory and Practice).*