Disputing Autonomy

Second-Order Desires and the Dynamics of Ascribing Autonomy

Joel Anderson
Utrecht University

Abstract
In this paper, I examine two versions of the so-called “hierarchical” approach to personal autonomy, based on the notion of “second-order desires”. My primary concern will be with the question of whether these approaches provide an adequate basis for understanding the dynamics of autonomy-ascription. I begin by distinguishing two versions of the hierarchical approach, each representing a different response to the oft-discussed “regress” objection. I then argue that both “structural hierarchicalism” (e.g., Frankfurt, Bratman) and “procedural hierarchicalism” (e.g., Dworkin, Christman, Mele) have difficulties accommodating the dynamics of how the attribution of autonomy to persons is claimed, disputed, and resolved. Although they differ in details, both shortcomings can be traced to viewing autonomy as a metaphysical rather than a normative, practical matter. I conclude by suggesting that these difficulties underscore the advantages of a more constructivist and pragmatist approach to autonomy.

Keywords: autonomy, second-order desires, hierarchical approach, procedural independence, Harry Frankfurt, Gerald Dworkin

Despite frequent claims that autonomy is overvalued, few concepts get as much use in the daily debates heard in hospitals, courtrooms, and social service

---

1 An earlier version of this paper was presented at Central Division meetings of the American Philosophical Association, where I benefitted greatly from the remarks of my commentator, John Christman. For further generous comments, criticisms, and suggestions regarding earlier drafts of this paper, I would also like to thank Kenneth Baynes, Jan Bransen, Bert van den Brink, Marcia Cavell, Randolph Clarke, Marilyn Friedman, Hinrich Fink-Eitel, Logi Gunnarson, Joseph Heath, Pauline Kleingeld, Thomas McCarthy, Robert Noggle, Henry Richardson, Paul Russell, Maureen Sie, Marc Slors, James Swindal, Candace Vogler, and Carl Wellman.
agencies. This is hardly surprising, given that most social interaction is premised on being able to attribute competence to others and that full human agency presupposes the self-trust that comes with a sense of one’s own autonomy. If this is so, however, then a theory of autonomy must also be measured by how well it is able to provide an adequate account of how autonomy is to be ascribed and how those ascriptions can be disputed and defended. In this essay, I examine the ability of recent hierarchicalist approaches to autonomy to provide such an account.

As I shall be using the term, “personal autonomy” refers to a capacity for self-governance, specifically, self-governance with regard to the choice of and commitment to personal projects, relationships, and ideals. On this understanding, being autonomous is a matter of being competent to guide oneself in such a way that what motivates one’s actions are the considerations by which one genuinely wants to be motivated, rather than (ordinarily, at least) by unconscious prejudices, compulsive obsessions, or the like.

As with other capacities, such as rationality or moral sensitivity, personal autonomy is a capacity that individuals develop to varying degrees in different contexts. Thus, although it is doubtless good policy and a matter of basic respect to start out from the assumption that others approximate the ideal of full capacity to some minimal extent, there are times and places where this everyday presumption of autonomy loses its appropriateness. In some cases at least – for example, those involving young children or the mentally incompetent – the important liberal prohibition on second-guessing others’ avowed desires may need to be suspended. Similarly, though perhaps more problematically, one may lose confidence in one’s own capacity for self-governance. Without claiming that we are constantly examining others’ competence to determine what they really want, one can still say that there is a sense in which our interactions with others are implicitly premised on their presumed competence.

2 For useful categorizations of conceptions of personal autonomy, see esp. Hill (1991, pp. 50-51) and Feinberg (1986, p.18), as well as Christman & Anderson (2005), Taylor (2005), and Mackenzie & Stoljar (2000).

3 This is not to say that people are necessarily responsible for the degree to which they attain autonomy, in the sense intended here. A person’s impaired autonomy is often due to his being denied the opportunity to develop it fully. Diana T. Meyers, for example, provides a compelling account of the way in which masculine and feminine socialization provide differential access to the acquisition of “autonomy competency” (Meyers 1989, pp. 50-51). In addition, as Marilyn Friedman emphasizes, a person may be clearly autonomous in some contexts, while less so in others (Friedman 1989).

4 This point has been made in a variety of contexts, using a variety of terms. Central examples are Peter Strawson on “personal reactive attitudes” (Strawson 1974), Robert Brandom on score-keeping practices regarding doentic status (Brandom 1994), Jürgen Habermas on the presup
My intent here is not to lay out an account of what precisely these minimal standards are, but rather to emphasize the significance, for a theory of autonomy, of the fact that autonomy is something ascribed and contested. In the practical contexts in which autonomy is at issue – for example, a patient’s request for assisted suicide or a murder defendant’s demand to represent himself – the relevance of autonomy-ascription is undeniable. My claim is simply that it is a desideratum of any theory of autonomy that its account of what autonomy entails provide at least the basis for an account of the pragmatics of how autonomy is to be ascribed and contested.

In what follows, I begin by distinguishing two versions of the so-called “hierarchical” approach to personal autonomy, based on the notion of “second-order desires” and developed most prominently by Harry Frankfurt and Gerald Dworkin. I then discuss these two approaches in detail, looking first at the “structural hierarchicalism” of Frankfurt (and Michael Bratman) and then at the “procedural hierarchicalism” developed by Dworkin (and Alfred Mele and John Christman). My primary concern will be with the question of whether these approaches provide an adequate basis for understanding the dynamics of autonomy-ascription. In the end, I suggest that the best prospects for doing so involve moving away from a metaphysical approach toward a more pragmatist and constructivist approach.

"Structuralist" and "Proceduralist" Responses to the Regress Objection

Hierarchical approaches start out from the compelling idea that autonomy is a matter of being guided not merely by the desires one happens to have but by the desires one reflectively endorses having. Thus, given a desire, say, to light a cigarette or refuse a life-prolonging operation, an autonomous agent is an agent who forms a second-order desire regarding whether the first-order desire should be effective in action. What distinguishes autonomous action and position of communicative competence (Habermas 1981), Harold Garfinkel’s investigations, via “breaching experiments,” of the everyday assumption of “accountability” (Garfinkel 1967), Donald Davidson’s “principle of charity” (Davidson 1984), and Daniel Dennett’s “intentional stance” (Dennett 1987).

In the following, I shall speak of “second-order desires” to refer to all such attitudes of reflective endorsement, identification, or rejection, ignoring Frankfurt’s oft-cited distinction between “second-order volitions” and “second-order desires”. Although the issue does not concern me here, the phrase “second-order volition” is rather awkward in any case, since it represents not a volition that has, as its object, a volition, but rather a desire with a particular content, namely, that the relevant first-order desire be effective. See Kusser (1989). For a different attempt to address the difficulties with this terminology term, see Stump (1988, pp. 400f.).
choice, on this view, is the presence of these second-order attitudes regarding one’s first-order desires.

Without further qualification, the apparent implication of this is that one’s second-order desires reveal what one “really” wants, that they are the site of one’s “true self.” This has provoked widespread criticism, particularly as leading to a regress (Watson 1975; Thalberg 1978; Friedman 1986). For the fact that one has second-order desires does not yet give one the necessary grounds for dispelling doubts about whether this represents what one really wants. These second-order desires might, for example, have been formed as the result of more or less subtle forms of brainwashing. Alternatively, one might be so internally divided about these matters that there would appear to be no clear fact of the matter as to what one wants at all. Critics contend that, on the hierarchical view, a third-order desire is required to authorize or integrate second-order desires in such cases. But again, third-order desires are prone to the same difficulties, and an infinite regress ensues. The point is quite general. If the authority or authenticity of any given desire of order “n” is established by appeal to a desire of order “n+1,” there seems to be no nonarbitrary stopping-point to the regress.

Defenders of the hierarchical approach now acknowledge that, as Frankfurt puts it, “The mere fact that one desire occupies a higher level than another in the hierarchy seems plainly insufficient to endow it with greater authority or with any constitutive legitimacy” (Frankfurt 1988d, p. 166f.).

According to hierarchicalists, however, all this shows is that something more is needed to stop the regress, and here two fundamentally different strategies emerge.

One strategy focuses on the structure of the reflecting agent’s will. On this view, which I shall call “structural hierarchicalism,” what makes a stop in the regress of reflection appropriate is the further condition that the identification with a desire be accompanied by a certain degree of integration and equilibrium within the volitional makeup of the person. Other hierarchicalists – such as Gerald Dworkin (Dworkin 1970, 1976, 1988b, 1988c, 1993), John Christman (Christman 1987, 1991, 1993, 1998), and Alfred Mele (Mele 1991, 1995) – take a different tack. They argue that the additional element needed is the requirement that one’s reflection occur under the appropriate conditions, namely, conditions free from coercion, misinformation, and the like. Call this

---

6 This is already clear in the 1976 essay “Identification and Externality” (Frankfurt 1988c, pp. 65f.) See also Frankfurt (1999c, p. 105) and Dworkin (1988b, p. 18).

7 The most important texts by Frankfurt in this regard are: Frankfurt (1988d, 1988e, 1988g) and Frankfurt (1999a, 1999b, 1999c, 1999d). Later works (Frankfurt 2004, 2006) basically continue the same line. Recently, this line has been further developed by Michael Bratman (Bratman 1999, 2006a, 2006b, 2006c).
“procedural hierarchicalism.” In their own ways, structural hierarchicalism and procedural hierarchicalism both seem to provide solutions to the regress problem just sketched: in order to determine whether one is choosing autonomously, one can look to either the coherence of one’s volitional makeup or the conditions under which one is reflecting.

Structural Hierarchicalism: Autonomy, Wholeheartedness, and Satisfaction

According to structural hierarchicalism (the approach suggested by Frankfurt), what is crucial for autonomy is that a person’s higher-order desires fit coherently into the structure of that person’s “will.” 8 On this view, what stops the regress of ever-higher-order desires – or rather, what makes a given stopping-point nonarbitrary – is a certain quality of the reflective agent. Frankfurt initially suggested that this was a matter of choosing “decisively” (Frankfurt 1988b, p. 21), and this is an idea that Bratman continues to emphasize (Bratman 1999). Frankfurt himself now uses the less voluntaristic expressions “wholeheartedness” and “satisfaction with oneself.” Frankfurt’s current view is clearly expressed in the following quote:

Hierarchical accounts […] do not presume […] that a person’s identification with some desire consists simply in the fact that he has a higher-order desire by which the first desire is endorsed. The endorsing higher-order desire must be, in addition, a desire with which the person is satisfied. (Frankfurt 1999c, p. 105)

Importantly, however, this satisfaction is neither an additional attitude nor a decision to quit reflecting. That would, after all, simply reintroduce the problem of a regress. Instead, satisfaction reflects “a state of the entire psychic system” (Frankfurt 1999c, p. 104), rather like being relaxed (Frankfurt 1999c, p. 105, note 16).

8 Although the term “structural hierarchicalism” is my invention, it is suggested by much of Frankfurt’s language. For example, in discussing the way in which “a person’s autonomy may be threatened even by his own desires,” Frankfurt formulates this as follows: “The determining conditions that are pertinent here are exclusively structural arrangements” (Frankfurt 1988d, p. 171). Although features of Frankfurt’s view suggest labeling it “integrationist” (Christman 1988, p. 113), the term “structural” better reflects the fact that, on Frankfurt’s view, the levels of endorsement remain distinct, even among the wholehearted. Finally, to avoid confusions regarding Frankfurt’s usage of the term “will,” it is important to note its decidedly un-Kantian character. Indeed, it is actually much closer to Hobbes’s usage (Hobbes 2000, p. 6, § 53). For an excellent discussion of some of the difficulties with Frankfurt’s notion of the will, see Fink-Eitel (1994).
The picture of the autonomous agent that emerges from this structural hierarchicalist account is the following. I am free from alien motives, and thus autonomous, when I wholeheartedly endorse the motivations for my actions, where “wholeheartedness” is understood as a state of the will that is untroubled by doubts or by a desire to change. For example, if someone cares wholeheartedly about choosing the career that will most please her parents, then she is in such a state that her pursuit of this end is not at odds with anything that is fundamental to her. Furthermore, insofar as she is genuinely in this state, there is no reason not to ascribe autonomy to her. Note that structural hierarchicalism is neutral with regard to the question of the rationality or origins of autonomous desires. One may be wholehearted about cruel or imprudent ends, and those ends may be the product of a bigoted upbringing. Frankfurt’s point is that this does not make it any less true that what one is governed by is one’s self.

In effect, then, structural hierarchicalism views wholeheartedness – together with the appropriate “active” capacity to form reflexive states (such as second-order desires) – as a sufficient condition for the autonomy of an agent’s current set of desires. If an agent wholeheartedly endorses her motives, she is a self-guiding agent free from alien desires. Wholeheartedness also represents a necessary condition for enjoying such autonomous status. For if a person is ambivalent or otherwise half-hearted, Frankfurt notes, there is nothing to which that person can be true, nothing internal by which he is being guided: “his ambivalence stands in the way of there being a certain truth about him at all” (Frankfurt 1999c, p. 100).

What basis, then, does structural hierarchicalism provide for approaching the dynamics of ascribing autonomy? Given the importance of wholeheartedness to autonomy, one would expect theorists following Frankfurt’s line to argue that assessing someone’s autonomy is largely a matter of gauging her wholeheartedness. And, indeed, Frankfurt does suggest viewing wholeheartedness as, “… a criterion for use in the design of ideals and programs of life, and generally in determining what to regard as important and to care about. What we care about should be, to the greatest extent possible, something we are...

---

9 In this regard, Frankfurt is quite close to Descartes, at least as he is understood by Frankfurt: “Descartes cares less about the correspondence of his beliefs to ‘reality’ than he does about their permanence and constancy. What he wishes above all to avoid is not error, in the sense of non-correspondence, but betrayal. What might be found out to be false is what he wishes to guard against. If a belief can confidently be expected to remain unshaken by any further inquiry, that is all the truth he cares to demand” (Frankfurt 1970, p. 180).

10 In the discussion after delivering “The Faintest Passion” at Northwestern University (May 7th 1992), Frankfurt was quite direct about this: “Stormtroopers,” he said, “can be wholehearted.” See also Frankfurt (1988b, p.19, note 6; 1988e, pp. 80-82).
able to care about wholeheartedly” (Frankfurt 1999c, p. 106). In particular, the coherence of the will found in wholeheartedness represents a criterion for what it means to be guided by oneself, that is, to be self-governing.

With regard to the issue of autonomy-ascription, the central question becomes how to determine whether a person is wholehearted. Because Frankfurt situates the stopping-point for the regress of higher-order reflection in a particular subjective state of the will (namely, wholeheartedness), his approach seems best suited to handling cases of self-ascription. Indeed, structural hierarchicalism handles them a bit too well. For it turns out that, in the first-person case, the question of whether or not one is wholehearted answers itself, and in a rather peculiar fashion. To be wholehearted is to have a sense that there is no need for further reflection. Thus, one doubts one’s wholeheartedness just in case one seriously wonders whether any of one’s desires are alien. But where there are no doubts of this sort, there is nothing else to be unsure about. If, by hypothesis, a person’s questioning comes to an end in a state of “self-satisfaction,” it is hard to see how doubts about whether to ascribe autonomy to oneself could ever really arise. Indeed, if one has no worries about one’s will – that is, if one is wholehearted – asking the further question of whether one is “really” wholehearted is, Frankfurt suggests, just so much foundationalist anxiety (Frankfurt 1999c, pp. 102-107; 1988d, pp. 167-169; 1988f, p. 155). For how could this psychological state itself be mistaken?

Frankfurt’s strategy here is subtle, but note that it amounts to sidestepping the issue of ascription by making the dubious assumption that the concept of wholeheartedness is self-applying. From the perspective of the deliberating agent, the question of one’s autonomy is supposed to answer itself by not being raised. But this makes it difficult to see how wholeheartedness or autonomy can serve as a basis for self-correction, of determining whether or not to trust one’s feeling of self-satisfaction. Suppose we took the same approach to “I’m fine; I don’t need my anti-psychotic medication.” For Frankfurt’s account to work, he has to view this as a fundamentally different issue from being wholehearted. But why?

In this regard, consider again Frankfurt’s parallel of self-satisfaction with being relaxed. It is unwise for insomniacs to try to fall asleep before they are fully relaxed. As many of us know, following this sound advice is hardly a matter of course, given how difficult it is to know when one is sufficiently relaxed. In the case of autonomy, a similar point holds. In order to follow Frankfurt’s principle of wholeheartedness – and thus to avoid not only frustration but also self-betrayal11 – the reflecting agent needs some way of telling when she has

been successful. And in this sense, there is no getting around the need for an account of how a person determines whether she is actually wholehearted, an account that Frankfurt himself seems to think unnecessary. Indeed, Frankfurt is able to formulate many of the puzzles and paradoxes that so fascinate him by placing seemingly familiar phenomena into a metaphysical space to which it is unclear how we could have epistemic knowledge.

In the interpersonal case, however, we can fairly easily imagine such an account, by considering the publicly available evidence for ascribing autonomy to others. Thus, in trying to determine whether a given choice of yours has been made autonomously, I must assess whether you are wholehearted about it. I can do this by looking at whether your actions agree with your words, whether your pursuit of your chosen ends is unhesitating, and so on. Whatever the details, however, it is essential to notice that it becomes possible to imagine a dispute arising between us: I sense that you are only half-hearted about some choice, and you disagree. If error is possible here – as in the parallel case of relaxation – there seems to be no reason to assume that it is necessarily I (and not you) who is in error. We have, then, a genuine dispute.

But this sits rather uneasily with the structural hierarchicalist understanding of autonomy as a reflexive state of the will, as a will that is satisfied with itself. The doubts of others cannot seem to get a foothold, on this account.\(^\text{12}\) In this sense, structural hierarchicalism encounters just the sort of difficulties plaguing the idea of a private language: without more of an account of how autonomy-ascription can be publicly contested and defended, we are left at a loss as to how to distinguish between being autonomous and thinking one is autonomous (Wittgenstein 1963).

Frankfurt could, of course, insist that these are all merely epistemological worries and that, as such, they leave the underlying ontological account unscathed. I have my doubts as to whether such a split between epistemology and ontology is actually coherent, but this would lead too far astray. Were one to concede the point, however, one would have to conclude that Frankfurt’s account is not intended to be of much use in analyzing ethical and political matters, where ascription issues are central. But Frankfurt clearly does have

\(^{12}\) This impression is reinforced by a passage in which Frankfurt draws a parallel between ending the regress of ever-higher-order desires and a regress of endlessly re-checking an arithmetic calculation (Frankfurt 1988d, pp. 167-169). What is striking about Frankfurt’s description of the latter case is that he construes it as a problem that occurs in the isolated chambers of the lone subject’s mind. The possibility of finding corroborating evidence or facing critiques from others is ignored from the outset in favor of a mentalistic criterion of Evidenz or self-certainty. In thinking about Frankfurt’s use of this analogy, I have benefited greatly from Candace Vogler’s unpublished discussion in Vogler (1992). See also my critiques in Anderson (1994, 2003).
practical implications in mind. Indeed, part of what makes Frankfurt’s recent writings so intriguing is their message that the chances of developing a strong, autonomous identity are being undermined by a culture so bent on ‘having it all’ that wholeheartedness becomes increasingly difficult to achieve. Thus if Frankfurt’s structural hierarchicalism is meant to show that volitional coherence is an ideal to which we should aspire, it seems fair to ask for more of an account of how it would be possible to determine whether we are making any progress. And here again, the issue of autonomy-ascription arises. The difficulty is not simply that Frankfurt has not yet provided such an account, but rather that his focus on the ontology of the will makes it hard to see how he could do so.

Procedural Hierarchicalism: Conditions of Unmanipulated Reflection

In many ways, the procedural hierarchicalism defended by Gerald Dworkin, John Christman, Alfred Mele, and others is better equipped to handle the dynamics of autonomy-ascription. Particularly for those working in contexts of applied ethics, there is much that is attractive about the apparently straightforwardly public criterion introduced by Dworkin: “if a person’s reflections have not been manipulated, coerced, and so forth and if the person does have the requisite identification [i.e., if the person has formed second-order desires] then they are, on my view, autonomous” (Dworkin 1988b, p. 20). This exclusive focus on the procedural conditions of critical reflection has, for many, the further advantage of being decidedly neutral with regard to the content of the choices that person has made or whether, in particular, she exhibits the “substantive independence” of a rugged individualist. According to Dworkin’s explicitly formal account, what matters for autonomy is that the process of critically reflecting on lower-order desires be characterized by the absence of interference. In this way, procedural hierarchicalism hopes to avoid making problematic, illiberal assumptions about which ways of life can count as autonomous.

Because procedural hierarchicalists focus on the conditions of critical reflection, a great deal turns on what the basis is for specifying the conditions of “procedural independence,” and for doing so in a “non ad hoc fashion” (Dworkin 1988b, p. 18). The basic idea is to distinguish between conditions that “interfere with the rationality of higher-order reflection and those that do not”

13 On this central theme of Frankfurt’s recent work – that an endless expansion of freedom actually jeopardizes freedom – see especially Frankfurt (1988a, p. ix; 1999a, pp. 112-16; 1982, pp. 319-21).
14 This neutrality is defended, in particular, in Dworkin (1988c). For important challenges to this neutrality see Oshana (2006) and Kristinsson (2000).
Disputing Autonomy

(Dworkin 1988a, p. 161). As it stands, however, this is still only preliminary, since what we are looking for is a way of distinguishing those influences that impair autonomous reflection from those that don’t. Dworkin himself grants that much work needs to be done, in cooperation with work in the empirical sciences, on what is to count as a condition of procedural independence, and one can imagine psychological data providing further insights into, for example, what serotonin levels or degrees of anxiety might impair reflective capacities in the relevant sense.

Several proceduralist theorists have subsequently contributed to this project of developing an account of the conditions under which critical reflection would count as authoritative. John Christman, for example, has argued that “the acts of critical reflection and identification with lower-order desires must not themselves be caused by Illegitimate External Influences” (Christman 1987, p. 291). Alfred Mele has developed a detailed account focusing on the circumstances under which one’s capacities for control over one’s mental life have been bypassed (Mele 1995, ch. 9). And John Martin Fischer and Mark Ravizza have developed an approach centered on the requirement that the mechanism responsible for a person’s process of critical reflection be at least moderately responsive to reasons (Fischer & Ravizza 1998; Fischer 2006).

One natural way to approach this task is to focus on relatively direct forms of external interference such as hypnotic suggestion, deception, and lobotomies, and the autonomy literature is replete with obvious cases in manipulation in which the source of the interference is clearly external to the agent: mad brain scientists, evil demons, hypnotists, and others who are clearly alien sources of interference. This emphasis on observable causal chains in the etiology of nonautonomous has the advantage of giving us a clear idea of what needs to be excluded.

Often enough, the doubts that arise about autonomy are focused on such visible contexts of interference. But this is clearly not the only way in which the rationality of higher-order reflection can be impaired. In particular, we must not overlook the various forms of mental or emotional instability that can also undermine one’s attempts to get clear on whether to endorse certain desires. For example, in the case of someone trying to decide during an anxiety attack whether to accept a tempting but risky job offer, the rationality of her reflection may be in doubt, but not as the result of any external interference (Young 1986; Stocker 1979). At one level, this is merely a call for further clarification with regard to what should count as the conditions of procedural independence. But the more one thinks these cases through, the clearer it becomes how complex these accounts must be.

Consider the following case, mentioned in an early article by Christman:
An elderly woman who is terminally ill and in great pain refuses a fairly minor operation that would prolong her life for another year. […] Imagine that her son has made various remarks concerning the expense of the hospital stay. The woman may, then, be caused to approve of her own desire to die as a direct result of this external pressure. If this were so, we would not regard her decision as fully autonomous… (Christman 1987)

Both the description and the conclusion are plausible enough. But getting to the conclusion is a very complicated matter, much more so than Christman indicates here. In particular, we need to have an account of the difference between the women being browbeaten by her son and her being convinced by him that it really would be best for her to “move on.” We need, in short, an account of what makes an influence alien and illegitimate.

At this point, proceduralist hierarchialism reaches a methodological fork in the road. With regard to any given way of spelling out the requisite conditions of authoritative critical reflection, one can take an ontological, realist, metaphysical approach or a neo-pragmatist, constructivist, post-metaphysical approach.15

On the former, “metaphysical” approach, the answer to the question of whether a given influence is alien or nonautonomous is a matter of objective fact, however difficult that fact may be to determine with any finality. Thus, for example, Mele provides an extensive and detailed analysis of the necessary and sufficient conditions for one’s authentically possessing a desire, that is, of the relevant pro-attitude not being compelled, where compulsion is understood largely in terms of having one’s decision-making capacities bypassed (Mele 1995, ch. 9). And, in a rather different but still decidedly metaphysical vein, Susan Wolf has provides an account of free or autonomous actions as in accordance with the “True and the Good” (Wolf 1990). Neither, however, sees it as a necessary part of their account to explain what the perspective is from which a determination is made, nor what the assertability conditions are for one person legitimately claiming that another person is (not) autonomous, nor what the effects are of ascribing to or withholding from someone the “is autonomous” predicate. As metaphysical approaches to the concept of autonomy, such accounts are concerned with what state of affairs the concept of autonomy depicts, represents, refers to, etc.

Approaches we can term postmetaphysical or neo-pragmatist, by contrast, focus on specifying the conditions under which it is appropriate to treat someone as being autonomous. Doing so is of a piece with understanding both the

---

15 My thinking about this rough, broad distinction between metaphysical and non-metaphysical approaches to concepts has been influence by a variety of thinkers, especially Williams (2004), Brandom (1994), Lance & White (2007), and Habermas (1992b). I develop these points further in Anderson (2007).
“upstream” considerations that support one’s attributing autonomy to someone and the “downstream” commitments that doing so entails. Thus, for example, Peter Strawson provides a parallel account of responsible agency in terms of the participant stance that we undertake and sustain in treating one another as fit targets of certain reactive attitudes (Strawson 1974). And Michael Smith and Philip Pettit analyze the form of quasi-autonomy they term “orthonomy” as tied to our ability to sustain a “conversational stance” with them (Pettit & Smith 1996). In all these cases, the analysis of the concept makes essential reference to a way in which individuals engage with each other from a second-personal standpoint.

I cannot pretend to provide a full assessment here of the relative merits of these two ways of approaching the concept of autonomy, in part because we still have a long ways to go in working out what a neo-pragmatist account of autonomy would look like. At this point, I wish merely to make two smaller points: that some established procedural hierarchicalists can best be understood as advocating a less metaphysical approach, and that moving in that direction has distinct advantages with regard to understanding the dynamics of disputes over someone’s autonomy.

Of the most influential procedural hierarchicalists, I would suggest that the approaches defended by both Christman and Dworkin can be understood as taking a post-metaphysical approach, at least some of the time. In one article, at least, Christman is actually interestingly ambivalent. He focuses on excluding influences that are “essentially external” to the agent (Christman 1987, pp. 289, 291). But on the same page he writes that his proposed focus on “Illegitimate External Influences” may be not be “ontological” (or in my language “metaphysical”): “the distinction between those external factors bearing on a person’s cognitive processes that do not subvert the autonomy of those processes, and those that do, is not based on strictly ontological differences between the two kinds of factors.” He then concludes, somewhat tentatively, that “The value of autonomous decision-making will turn out to be just as ‘conventional’ as the specification of the phenomenon itself, which might bring into question some of our deeper intuitions concerning those processes” (Christman 1987).

Christman’s negative or at least skeptical point is complemented by one reading, at least, of Dworkin’s approach. He is explicit that the central distinction between cases in which autonomy is or is not violated “… is not a metaphysical distinction but a practical one” (Dworkin 1988a, p. 161). He elaborates this elsewhere as follows:
... I do not think it possible with any moderately complex philosophical concept to specify necessary and sufficient conditions without draining the concept of the very complexity that enables it to perform its theoretical role. Autonomy is a term of art introduced by a theorist in an attempt to make sense of a tangled net of intuitions, conceptual and empirical issues, and normative claims. What one needs, therefore, is a study of how the term is connected with other notions, what role it plays in justifying various normative claims, how the notion is supposed to ground ascriptions of value, and so on — in short, a theory. (Dworkin 1988b, p. 7)

Dworkin is here defending a decidedly constructivist, pragmatic approach. (This anti-metaphysical approach is, incidentally, one of the reasons why it is a mistake to lump him together with Harry Frankfurt, who is very explicitly interested in developing an ontology of the will.) It may not quite be the more radically anti-metaphysical stance approach, but it keeps the door open to such a view.

At this point, we can ask whether this is an advance or not, relative to more metaphysical approaches to procedural hierarchicalism, such as Mele’s. One way of answering this question would be in terms of the overall merits of metaphysical and pragmatist approaches, with an eye to how it has worked out in other domains, such as epistemology or philosophy of mind. But such an assessment is clearly beyond the scope of this short essay. Suppose, however, that neither methodology could be eliminated on those grounds alone. We could also assess the relative merits in terms of the ability of each to fit with what we know about the real dynamics of how autonomy attributions are made and disputed.

If we did this, defenders of metaphysical approaches could clearly point to certain advantages. For example, metaphysical approaches seem to explain why we experience disputes over whether someone is autonomous as disputes about something that we could get wrong, suggesting that, if disputes over someone’s autonomy are not simply to be resolved by the exercise of power, there must be some truth of the matter of what autonomy is. Relatedly, it might be thought that only a metaphysical account allows us to make sense of the idea that we might collectively be profoundly mistaken in our understanding of what autonomy really is, that is, that our intuitions might largely converge, but on a distorted view of matters. Arguably, this has been the case in the past with regard to women or racial, ethnic, and religious minorities.

On the other hand, if these points can be accommodated — a point I will return to in a moment — the neo-pragmatist approach has a decided advantage in showing how central it must to be understanding how we dispute autonomy that the objections people make to the attribution or withholding of the status
of being autonomous are *normative objections to being treated a certain way*. For what I dispute, when I dispute your refusal to see me as autonomous, is not in the first instance a representational a matter. Rather, it is a dispute over what my status is, for example, with regard to my refusing life-saving treatment. Metaphysical approaches tend to make it hard to accommodate these dynamics of how autonomy gets disputed by approaching “is autonomous” as a predicate whose extension is fixed definitionally, in terms of a particular set of necessary and sufficient conditions. We could then treat it as a separate issue whether or not it is a good thing to be autonomous. That is, we could separate the question of *what it is to be autonomous* from the question of *what the practical or normative implications are of being autonomous*. But saying “You’re not autonomous, but that doesn’t mean you cannot give informed consent” is a bit like saying “You’re ugly, but I mean that in a purely descriptive sense.” Indeed, as Mark Lance and Heath White have recently argued, in favor of neo-pragmatist, “stance” approaches to personhood (and against “metaphysical” approaches), “The danger facing metaphysical approaches is that they will miss whatever non-cognitive aspects there are to, say, personhood. This is true even if they give accurate necessary and sufficient conditions for personhood, which make no reference to appropriateness.” (Lance & White 2007) What I take the (in my view, unfortunate) talk of “non-cognitive aspects” to refer to are just these practical and normative implications, implications paradigmatically involving shifts in one’s deontic or normative *status*. My asserting that you are autonomous entails committing myself to your having a certain status, a status that licenses various ways of treating you and prohibits others. This normative dimension is not separate from the question of whether the predicate is appropriate. It is, crucially, what disputes over the attribution of autonomy are largely about, and neo-pragmatist approaches are able to avoid getting sidetracked by issues of “what autonomy is” and focus instead on the complex and contested matter of how we ought to treat one another.

Of course, the post-metaphysical, pragmatist approach is going to be plausible only if it can provide a satisfactory alternative to the points made earlier, in defense of metaphysical approaches, about the possibility of deep error. There are several approaches that can be made here. I shall limit myself to the following point. The neo-pragmatist position is that metaphysicians do not have a monopoly on fallibilism. On the contrary, neo-pragmatist and stance-dependent accounts of autonomy allow plenty of room for error. The difference is that they locate that error differently: what we get wrong is not a representational issue but a practical, normative issue. There may well be an error, but not an error with regard to “what autonomy is”; rather, it is in the first instance an error in how we treat people. And it is simply begging the question
to insist that the only way we can understand errors in how we treat people is in metaphysical terms.

Conclusion

The points just made also apply to various ways in which we could approach structural hierarchicalism, that is, with regard to concerns with the role that wholeheartedness and self-satisfaction play in appropriately stopping the regress of critical reflection. Indeed, it starts to look as though the fundamental difficulties with most standard forms of hierarchicalism lie in their attachment to a metaphysical approach. Such approaches seem most promising when autonomy is presented as a private, internal matter of the will (as in the case of structural hierarchicalism) or as a matter of clear-cut objective criteria (as in the case of some defenders of procedural hierarchicalism). But many of cases for which we want an account of autonomy to help us think matters through are cases where we have doubts about our subjective sense of how things stand or where the only way to determine whether conditions of procedural independence obtain is by considering whether the reasons that the individual gives are cogent. But if this is so, then in any situation of trying to assess autonomy, one can no longer simply assume that the person who is officially placed in the position of applying the principle is not impaired in her judgment or applying the wrong principle. Judges, doctors, and psychiatrists have neither privileged access to good reasons nor any guaranteed ability to recognize good reasons. The possibility that one is operating under conditions that are not actually those of procedural independence applies symmetrically to the person whose autonomy is in question and those who are trying to assess her autonomy. This is our situation with regard to the disputes that can ensue at this point: autonomy-ascription is not simply a matter of applying a set of objective criteria, but is rather premised on the mutual and defeasible assumption on the part of interlocutors that each is competent to offer and evaluate good reasons and that both the criteria and their application are open to dispute.

Once it is recognized that any adequate account of autonomy will have to acknowledge that no one’s understanding – neither self’s nor other’s – of what is to count as grounds for ascribing autonomy is immune from being contested, in principle, within a public, dialogical process, it makes even more sense to view autonomy not so much as an objective metaphysical property of persons rather than as a socially situated status that one comes to enjoy in virtue of securing, as an ongoing, everyday achievement, others’ recognition of one’s autonomous agency. On such a nonmetaphysical view, the question of whether or not one is
an agent of autonomous deliberation can be answered only in the give-and-take of providing, challenging, and defending reasons for one’s position.\(^{16}\)

Such a neo-pragmatist, nonmetaphysical approach is not without complications, particularly regarding how to acknowledge the ever-present possibility that those who are deemed incompetent do in fact deserve to be considered autonomous. Nonetheless, it holds out the prospect of avoiding the difficulties we encountered with both structural hierarchicalism and procedural hierarchicalism. It takes autonomy out of the domain of metaphysical speculation and puts it in the realm of publicly redeemable claims and to political conceptions of how autonomy relates to our other values and normative concepts. And it does so without either relying on a substantive account of rationality or privileging any perspective as having special insight into what counts as the proper procedural conditions, since there is no reason to think – in advance of the actual process of asserting, challenging, and defending positions – that anyone’s understanding or application of the grounds for ascribing autonomy should be beyond dispute. For we can never know in advance whether the individual is failing to meet the criteria or the dominant criteria are failing to meet the patient.

Joel Anderson
Department of Philosophy
Utrecht University, the Netherlands
Joel.Anderson@phil.uu.nl

Works Cited


\(^{16}\) The outlines of the sort of approach envisioned here are suggested by (inter alia): Peter Strawson’s discussion of the way in which “personal reactive attitudes” attribute morally responsible agency (Strawson 1974); Jürgen Habermas’s conception of identity as dialogically constituted through mutual critique and recognition (Habermas 1992a); Harold Garfinkel’s ethnomethodological research on “accountability” and gender identity as “ongoing achievements” (Garfinkel 1967); and Axel Honneth’s analysis of autonomy in terms of “struggles for recognition” (Honneth 1995).


— (1988e) ‘The Importance of What We Care About’. In: The Importance of What We Care About. New York: Cambridge University Press, pp. 80-94.


