The Uses of Hierarchy: Autonomy and Valuing

Neil Roughley

Abstract

Autonomy and valuing are two significant practical phenomena that have been analysed in terms of higher-order wanting. I argue that reference to higher-order capacities is indeed required to make sense of both concepts, but also that such analyses need a more differentiated understanding of “wanting to want” than has hitherto been proposed. Central for autonomy is the instantiation of four types of optative relationship by an accountable agent under conditions of rationality. Valuing requires the disposition to instantiate only one of the relevant structures. Clarity on this allows an analysis of the precise relationships between the two phenomena.

1. Introduction

1. Autonomy and valuing are two significant practical phenomena that have been analysed in terms of higher-order wanting (Dworkin 1976, 1988, 1989, Lewis 1989). The use of the hierarchical apparatus for these purposes has also been contested (Thalberg 1989; Mele 1993; Blackburn 1998, pp. 66f.). In this article, I argue that reference to higher-order capacities is indeed required to make sense of both concepts, but also that such analyses need a more differentiated understanding of “wanting to want” than has hitherto been proposed. Only armed with the relevant distinctions can we proceed to clarify the precise relationships between valuing and autonomy.

2. Since Kant, the term “autonomy” has been used to stand for a “positive” notion of free will capable of founding moral responsibility. I shall argue that we would be well advised to sever such strong connections and instead to understand autonomy as a sophisticated way of relating to oneself that merely presupposes whatever concept of free will can be shown to be necessary for accountability. The reflexive relationship picked out by the metaphor of self-rule turns out to involve a cluster of four analogous, but distinct forms of higher-order wants, where these are formed under conditions of deliberative rationality and prove to be causally effective. Such efficacy constitutes a form of inner freedom accessible to accountable agents. As we tend to place a high value on the possession of these character structures, autonomy thus understood is accurately classified as a (Humean) virtue, one realisable in varying degrees. Moreover, once the topic of autonomy is thus unburdened metaphysically, the problem of the authority of highest-order wanting, i.e. of the “self” doing the governing, loses its recalcitrance.
3. Valuing, although a plausible candidate for the subject of the governing relation, is unable to fulfill this role; its analysis, like that of autonomy, requires reference to a higher-order capacity. Nevertheless, valuing does not require that capacity's realization. Moreover, the relevant capacity concerns only one of the four structures whose realization autonomy requires. Autonomy is exercised whenever one of the relevant structures is instantiated by an accountable, reflective agent, where the relevant highest-order want is reflectively uncontested and itself not irresistible. Certainly, autonomy is a capacity we value because it enables us to realize our values. However, it may also be coherently valued because it permits a form of freedom from our values.

2. Concepts of Autonomy

2.1. A Metaphysics of the Person

1. Applied to persons, "self-government" is a metaphor. In talk of personal autonomy, neither the governing relation nor its relata — its subject and its object — are easily identified. Kant, the philosopher responsible for applying the concept to human agents, saw it as such a problem. For Kant, the governing relation is constituted by the action-guiding efficacy of a principle or "law", the object of that relation is the person, in as far as she is an inhabitant of the sensible world and the relation's subject is the person, in as far as she is an inhabitant of the intelligible world (1785, p. 453). Where autonomy is given, the "I", that which in the person is "pure activity" (1783, p. 451), imposes on the empirical agent principles of action characterised by independence from any other motive of which she may happen to be the bearer. Action according to any such empirically given motives is "heteronomous", as such motives will all have been caused by processes outside the control of the person. It famously turns out that the only principles that satisfy the criterion of independence are principles acceptable to all fully rational beings — what Kant calls "universal laws" (1785, p. 440). The freedom thus gained from one's own empirical motivation is the flip side of the positive freedom that consists in imposing those laws "oneself" (1785, p. 446f). For Kant, freedom and autonomy are "reciprocal concepts" ("Wechselbeziehungen"), which name two sides of the same coin (1785, p. 450).

2. If the metaphysics of two worlds were the only way one could make sense of the notion of self-government, then it would be wise simply to drop it. Even if one accepts the epistemological necessity of positing a realm of "things in themselves", there is little to be gained by making it the locus of spontaneity. Although, according to Kant, causality is only operative within the realm of appearances, the governing relation can only be given if there is a form of causality that operates across the border between the two worlds. On the other hand, the desire to act according to universally acceptable principles is as empirical a motive as any other and therefore equally subject to genesis by causal processes over which the person may have had no control.

2.2. Patient Autonomy: Generation of Context-Specific Normativity

1. If there is a property appropriately described as self-rule, then we are going to need different conceptual tools to capture it. Certainly, there are contexts in which people talk of "autonomy" without for one moment conceiving the persons thus described as divided between two realms. One such context is the discussion of informed consent in relationships between doctors and patients. Beauchamp and Faden (1986, p. 235f.), for instance, suggest criteria for a notion of autonomous action, such that, where consent to medical treatment satisfies those criteria, the consent generates normative force, "binding" the consenter. In their conception, autonomous consent to a form of treatment is intentional authorisation of that treatment under conditions of "substantial understanding" and "substantial non-control" by others (ibid. p. 278). The concept thus circumscribed locates thresholds on the two axes of "understanding" and "control", but without being able to mark the precise point at which such "substantiality" begins, i.e. beyond which consent takes on its normativity-generating function.

2. The task of delineating a concept of "autonomy" for this particular context is constrained by the institutional and economic structures of specific forms of medical practice. Its status is thus far removed from that of the metaphysical concept for which Kant argues. Two points of difference are worth emphasising. Firstly, whereas Kantian autonomy is an all-or-nothing affair, contradictorily related to heteronomy, a notion of autonomous action crafted for the purpose of medical ethics is located on a continuum beyond a threshold which, under changed institutional conditions, could justifiably be shifted. There is no point in crafting concepts for specific purposes if they turn out to be maladapted to the conditions of their application. Secondly, the two conceptualisations are marked by divergent levels of interest in the reflexivity that is at the heart of the term's etymology. For Kant, persons are able to govern themselves because of the internal split that allows their intelligible part to enter into a relationship with their empirical part. In contrast, a suggestion such as that of Beauchamp and Faden gives the term's reflexive component a non-reflexive reading. That a person governs herself is taken to mean nothing more than that she acts, where certain adverse epistemic and volitional conditions are absent. What the person "governs" are her actions, rather than any particular feature of herself. The construction and the criteria are basically those used by Aristotle to classify actions as voluntary ("hekousion") (NE 1109b30-1111b3).

2.3. Autonomy and Freedom

1. The sharpening of those Aristotelian criteria has often, correctly, been seen as the task of a theory of freedom. According to an influential suggestion within political theory (Berlin 1969, pp. 122ff.), positive and negative freedom are two logically separate concepts. This claim has in turn been taken up in contemporary discussions of personal autonomy and used to drive a conceptual wedge between autonomy as positive freedom and its cousin, mere negative freedom. For instance, Gerald Dworkin has claimed that autonomy involves a restriction of
someone's autonomy, but not of their freedom, whereas an act of self-binding, such as that of Odysseus, involves an increase in the subject's autonomy and yet a decrease in their freedom (Dworkin 1989, p. 59f.; 1988, pp. 14f.). In the first case, although no barriers are placed in the agent's way, she is prevented from determining her actions herself. In the second case, the person takes active control of the events he is involved in precisely by hindering himself in his potential actions.

2. However, the opposition between freedom and autonomy grounds in a superficial analysis. Freedom, as Hobbes (1651, p. 136) saw, is generically a matter of non-obstruction. This does not constitute a specifically "negative" notion of freedom for the simple reason that talk of "obstruction" makes no sense unless there is some factor that can be obstructed (cf. Feinberg 1980, pp. 3ff.). Call this necessary component in the structure of freedom the obstructable factor (OF). The OF is what makes certain options significant and others irrelevant. Thus, any concept of freedom has to furnish criteria of what is to count as an OF: Concepts or theories of freedom developed for particular contexts will go on to specify what is to count as its obstruction. The OF of a free falling stone is the effect of gravitation on its mass; obstructions can be just about anything that gets in its way. A free-range hen is largely unobstructed in its natural striving to move about; and enclosures smaller than a certain size count as obstructions.

3. Turning to freedom of the person: for Hobbes the relevant OF is what a person wants. However, according to Hobbes, not every obstacle to the realisation of a person's want counts as a restriction on freedom. Hobbes stipulates that only external factors qualify as relevant "impediments", whereas internal factors, whether permanent or temporary, can merely diminish a person's "power". Thus, if someone is ill and is for that reason unable to do certain things they want to do, they do not as a result suffer a decrease in freedom. In fact, if one takes Hobbes' suggestion seriously, someone who is suffering from an illness may well find their freedom increasing, as the number of obstacles to what they have the power to do diminishes.

4. Hobbes' stipulation is clearly arbitrary. There is no conceptual reason for drawing the line between OF and obstacles in terms of the distinction between "internal" and "external"; interpreted with reference to a person's body. If, in order to prevent someone escaping, I inject him with some drug or virus instead of locking him in a cell, I can deprive him of his freedom of movement to precisely the same extent. The same means can also be applied in order to deprive someone of their freedom of thought or their freedom to develop or realise wants. Moreover, the question of whether the impediment in question is the result of deliberate human intervention or unforeseeable causal chains is irrelevant for its character as a diminisher of freedom.

5. Finally, the concept of freedom as non-obstruction leaves it open whether the absent obstruction itself consists of the presence or absence of some state of affairs. If someone wants to go from A to B, his freedom to do so can be negated either by a wall or by a hole in the ground. Similarly for internal obstacles: the injection of a drug can equally restrict a person's freedom whether it works by causing temporary blindness or unbearable pains whenever the person moves. This also applies to those internal states we call beliefs: you can prevent a person from achieving something by either feeding him with false beliefs or keeping him in ignorance.

6. According to this structural conception of freedom, deception can be an impingement on a person's liberty as significant as certain forms of physical obstruction. This is not an area where talk of "self-rule" names considerations uncovered by the concept of freedom. Phenomena of self-binding are structurally complicated, in that they involve the self-imposition of restrictions in order to overcome actual or anticipated impediments to getting what one wants. Self-binding is thus a deliberate restriction of one's freedom in certain dimensions in order to increase freedom in respect to aims that are in some way given greater weight.

7. The plausibility of the claim that the personal property thus picked out is not a kind of freedom grounds in the idea that binding oneself can involve a net loss of freedom, where the level of freedom is calculated in terms of the number of wants whose realisation is unimpeded. But there is nothing special about this among the varying spheres of freedom. Freedom to do a may only be attained at the cost of freedom to do b, c and d, independently of whether getting to do a results from taking measures to prevent one's desires for b, c and d influencing one's behaviour. One may simply be standing at the cross-roads, one fork of which leads to a and the other leads to b, c and d, where geographical and temporal factors combine to allow only either a or the conjunction of the other three options. Once one has gone a certain way down the road to a, one is no longer free to go to b, c and d. In this respect there is nothing special about self-binding.

8. If it is correct that autonomy is a species of freedom, then there is every reason to suppose that, like the generic property, its reflexive form comes in grades, so that people can be more or less autonomous, just as they and other entities can to varying extents be free. When certain people are identified as unqualifiedly autonomous, what is meant is that they possess the relevant property to an extent that carries them across a threshold taken as a standard in that context.

2.4. Accountability

1. As both a gradable and a reflexive property, autonomy would be unsuitable as a foundation for the accountability of persons, a role it was assigned by Kant (cf. Wolf 1990, pp. 10ff.). What accountability does presuppose is rather a fundamental form of freedom that is neither gradable nor reflexive. What constitutes the relevant form of freedom is, of course, highly disputed. Kant's belief that it involves "the capacity to initiate a series of events" oneself, where that capacity is the "unconditioned condition" of voluntary action (1781, p. 375), is shared by some present-day agent causationists, although, unlike Kant, they see the capacity as instantiated by empirical human persons. Others require less. Whatever conception one develops here, it cannot plausibly ground accountability unless it
instantiates the structure of some unobstructed OF. For instance, Fischer and Ravizza's conception of a "reason-responsive mechanism" (1998, p. 69ff.) is only going to be a candidate for grounding accountability if it is to some significant extent unimpeded in its functioning and in its influence on the action of its bearer. And the mere fact that certain people find themselves manifesting the reactive attitudes towards someone is decidedly insufficient, unless they are able to point to some unimpeded factor at work in the objects of those attitudes that justifies them (cf. Bennett 1980, p. 18).

2. Whatever conception one inserts here, its realisation is going to be a precondi-
tion for a person counting as autonomous. Autonomy is a property possessible in varying degrees by accountable persons. Bearers of the freedom to act responsi-
ably are candidates for the sophisticated property, or as I shall argue: property clus-
ter that constitutes a special kind of inner freedom. This is a property the realisa-
tion of which tends to be highly valued in our culture.

2.5. A (Humean) Virtue

1. I want, then, to suggest that the most appropriate use of the term "autonomy" is to pick out certain structures by means of which accountable agents relate to themselves. For the bearer of the relevant reflexive structures to count as autonomous, those structures must be effective in bringing about actions or omissions of the person. The reason why we often use the term to pick out these causally effective reflexive structures is that we tend to consider them of consider-
able value. In fact, being possessed of these structures fits the bill very nicely of what, according to Hume, constitutes a virtue (1739/40, p. 591). Autonomy thus understood is a structural property whose exercise tends to enable its bearer to act, or refrain from acting, in ways conducive to her well-being. It also tends to contribute to her self-esteem. Moreover, the relevant characteristic is also gener-
ally both useful and pleasant for such a person's partners in interaction: autonomous persons can by and large be relied upon and we tend to admire them.

2. Autonomy thus understood is no candidate for a number of roles in which the term, along with certain components of the concept, has repeatedly been enlisted. It can explain neither freedom (of the will or otherwise) nor accountability (moral or otherwise). It is not a necessary condition of consent capable of binding its bearer. And it is certainly not the basis for the generation of moral or political rights (a further role of Kantian autonomy). Its status as a virtue in the Humean sense means that it may well only be strongly valued in historically and culturally restricted contexts. The extent to which it is and has been present in other cultures is an empirical question. It certainly seems highly plausible that it is a characteris-
tic all undamaged humans have the capacity to develop. Note finally that the characterisation of autonomy as a Humean virtue is neutral relative to questions of its moral status. Humean virtues are not necessarily moral virtues (1739/40, pp. 606ff.) and there is every reason to suppose that autonomy as it is understood here can be enlisted in pursuing projects of dubious moral status.

3. A Reflexive Virtue in Four Variants

3.1. Frankfurtian Hierarchy: Realisation-Oriented Wanting

1. According to a well-known suggestion, the reflexivity essential to self-rule can be explained in terms of the attitudinal relationships between the person's wants, or what I shall call optative relations. The claim that higher order wanting is at least in part constitutive of autonomy was first explicitly advanced by Gerald Dworkin (1976), although the more widely discussed use of the theory was put forward by Harry Frankfurt (1971) as an explication of freedom of the will. In what follows I shall argue that hierarchical relationships between wants are indeed at the heart of the virtue of autonomy. However, the presupposition of the concept's applicability, that the bearer of the attitudinal relationships is account-
able and thus in some relevant sense free, cannot be satisfied by any such purely internalistic specifications. This is primarily because what is often called "freedom of the will", that is, a person's freedom in wanting, is to a significant extent a his-
torical matter: it depends on the conditions under which the relevant wants were formed. This entails that the OF that is relevant to the question of free will is not entirely explicable in optative terms, but is at least in part a part of proto-optative dispositions to want formation. A theory of free will is therefore to a significant extent a theory of what conditions count as impediments to the humanly normal processes of want development. The bearer of wants that are the spec-
ific products of hypothesises, neurological manipulation or systematic indoctrination is going to be unaccountable for actions caused by those wants and is thus, as regards these actions, no candidate for autonomy.

2. In the article that effectively launched the modern discussion of optative hier-
archy, Frankfurt focuses on what I think we should see as one particular form of higher-order wanting: namely wanting that the content of a particular first-order want be realised. For Frankfurt, being the bearer of second-order realisation-orien-
ted wants, what he calls "second-order volitions", is the essence of being a per-
son. A person is, according to his conception, essentially a being capable of prac-
tical rationality, that is, of establishing a reflexive relationship to her motivational attitudes and on that basis taking on a further optative stand towards them. A spe-
cific form of freedom is given when that optative stand is effective in mastering the motivational force required for the person's first-order want to cause an action accordingly describable in terms of the want's content. We justifiably apply the con-

---

1  I use the adjective "optative" to mean want-related, where "want" stands for the generic attitude verbally expressed by "Let it be the case that p" and characterised by a world-to-content direction of fit. I prefer "optative" to "volitional" (generally used by Frankfurt) because willing entails addi-

2  Describing the wants as "the exchise products" of such processes is intended to exclude two kinds of case which would otherwise invalidate such a historical condition: either the person is in part the bearer of the relevant wants because of not exercising the occurrent capacity to rid himself of them (cf. §4.3.4) or she has in the past deliberately arranged for the relevant historical process to take place.
cept of freedom because there is a lack, or at least an insufficient level, of obstacles to the second-order attitude effectively contributing to the realization of the ground-floor want. Largely because of the historical problems mentioned above, the causal efficacy of this structure is, pace Frankfurt, insufficient for freedom of the will. However, once such defeating conditions in the formation of the higher-order want are excluded, then someone who is host to such an optative structure is a candidate for a form of reflexive freedom sensibly termed “autonomy”.

3. The core idea of autonomy as I am understanding it is that of a relationship between the wants of an accountable person, a relationship established in the light of reflection and characterised by causal efficacy. A person “governs herself” in this sense where she comes to grips with her own motivationally relevant attitudes in this both reflective and reflexive manner. As such, “self-rule” requires a level of self-transparency in a person’s motivational processes that is no doubt significantly restricted in its empirical instantiation. Were we attempting to pin down the basis for responsibility or the attribution of rights, that would be a problem. These latter concepts require foundation in non-reflexive properties of their bearers. In contrast, the virtue of autonomy is plausibly a reflexive property — or, as shall argue: a cluster of reflexive properties — that has come to be particularly valued in a specific socio-cultural context. For this reason it is also to be distinguished from the Platonic and Aristotelian virtue of temperance (“sophrosyne”) (Plato, Pol. 442ac-d; NE 1117b23-1119b21) and the Aristotelian sub-virtuous characteristic of “enkratiea” (NE 11145a15f.), neither of which requires the kind of reflexivity essential to autonomy and often diagnosed as peculiarly modern. Although there are overlaps, particularly between “enkratiea” and autonomy, it makes good sense to mark a distinction, as doing so makes it possible to discuss the relative merits of these varying forms of control.

4. Realisation-oriented second-order wants can be positive or negative in their polarity, mediated or immediate in their efficacy. They are most noticeable in their

negative form, for instance where the bearer of want to smoke wants that first-order want not to be realised or where someone suffering from an outbreak of jealousy struggles to prevent her jealousy from determining her action. Such struggles can have an immediate optative form or else be mediated by strategies. The strategies can in turn be either external, such as dipping your cigarettes in mustard, or internal, such as focusing your thoughts on other matters.

5. Note that the employment of these strategies is no conclusive proof that the person is engaging in the reflexive relationship required for autonomy. Certain people might have been taught to engage in focusing techniques whenever they find their mind wandering from unpleasant tasks and they might do this without being aware that the technique is one which functions by working against countervailing wants. For instance, a child may employ such techniques because he has been told to do so and because he has learnt that he is rewarded when he does so (cf. Mele 1995, 67). The lack of self-transparency in such a case would exclude even the successful use of the technique from qualifying its bearer as autonomous.

6. Such a case is in turn to be distinguished from cases of dispositional second-order wanting, where particular kinds of second-order wants have been consciously formed and are able to be activated and effective in contexts in which they no longer require conscious mediation. Take Morgan, who has sworn not to let his chronic desire to insult his mother-in-law get the better of him again. If when he finds the desire to do so welling up in him, he – either directly or by biting his tongue – manages to overcome it, the efficacy of his negative realisation-oriented desire need not depend on him having a conscious thought with that content. There is good empirical evidence for the claim that, where someone has managed to install a conditional optative disposition in himself by earlier conscious thought, it can, when the antecedent appears to be satisfied, function effectively below the level of consciousness (cf. Bargh/Barnsdall 1996, pp. 460ff.). The inauguration of the dispositional want through conscious thought and its continued conscious accessibility mean that the transparency condition is nevertheless fulfilled.

7. It should be noted that a person’s capacity to take an effective negative realisation-oriented stand on some first-order want can serve as a defeating condition with respect to those historical conditions, such as indoctrination, that would otherwise exclude an agent from counting as accountable in acting on that want. If indoctrination has led a person to want something, but she is nevertheless able to restrain herself from acting on it, then her action to realise the first-order want is something for which she is accountable.

3.2. Acquisition- and Retention-Oriented Wanting

1. The reflexive optative relationship required for the virtue of autonomy is not necessarily established in terms of realisation-orientation. Frankfurt (1971, pp. 14f.) pointed out that a person can also want simply to “have”, but not to realise a ground-floor pro-attitude. Although Frankfurt referred to such higher-order wants as “truncated”, they are in fact semantically richer than realisation-oriented
wants. The latter are simply reapplications of the wanting's mode to a first-order want. As the point of a want is that its content be realised, its mode can be expressed by "Let it be the case that \( p \) or "Let \( p \) be realised". Realisation-oriented wants thus, at least in standard cases, have the structure "Let it be the case that (Let it be the case that \( p \))". However, where the content of a higher-order want includes a presentation of its bearer as the bearer of the lower-level attitude, that content is then about the bearer in this role and is not just directed at the want. Where this is the case, the higher-order attitude has the structure "Let it be the case that I have a want with the content (Let it be the case that \( p \))". It turns out that there are three ways of wanting to "have" a want of a lower order. What distinguishes these ways of higher-order wanting are different specifications of the sense of "have".

2. A first way of wanting to "have" a lower-order pro-attitude is to want to acquire it. For example, on seeing a film about Mother Theresa, Ben develops the want to be the sort of person she is presented as having been, i.e. Ben finds himself wanting to have the motivation to commit his life to helping others. Note that there is no reason why such an acquisition-oriented want need be successful: however much he may want to develop altruistic wants, he may remain unable to be moved by the concerns of others. Acquisition-oriented wanting can also be negative in character. Little Solomon has heard about the health risks of smoking, but fears that, if he continues hanging around with the kids on the block, he will end up wanting to smoke too.

3. A second sense in which someone may want to "have" a pro-attitude is made clear by the following thought: whether or not a particular first-order want of a person has been developed autonomously, it is conceivable that she can at any time form a higher-order want whose content specifies that she either retain or discard that first-order want. Grown Solomon might, after having been unsuccessful in preventing the genesis of a desire for nicotine, forget about his prior aversion to the desire and just get on with first-order craving. One day, however, on seeing a film showing the gory details of the inside of a smoker's lung, he may be prompted to re-think. The conclusion of such a re-think, if it is determinate, will be either positively or negatively retention-oriented. That is, he can come to the conclusion that, however he came by the want in the first place, either it is one that he has to be rid of or else one that, in spite of the health risks it brings, he endorses as bringing a form of relaxation he wouldn't want to live without. Perhaps some retention-oriented wants can take direct effect. Someone who first reacts to some situation by wishing a rival harm may, on thinking through the case, conclude that the desire is childish and optatively distance herself from it in such a way that it dissolves. Mediated efficacy of such wants is often via repeated abstention from the satisfaction of the first-order attitude, an effect itself perhaps the result of a negative realisation-oriented want. The mediating strategy employed may also be mental, as when someone wants to retain his commitment to a lying relationship and therefore deliberately recalls all the pleasant experiences it has involved.

4. It should also be noted here that the capacity for effective negative retention-oriented wanting can fulfil the same function as that claimed for realisation-orien-
For this reason, he develops a higher-order want that his dispositional want remain unactivated.

3. Compare Emma. She is a naturally empathetic person. Unlike Solomon, she endorses the retention of her first-order want—to help people and animals in need. Nevertheless, she finds that it is sometimes activated in the most unfortunate situations, a pattern that has cost her a number of jobs. Emma therefore forms a negative activation-oriented want in order to deal with situations where she is supposed to be getting on with her work, independently of whatever suffering may be going on around her.

4. The strategies adopted to prevent want activation are by and large the same as those adopted to prevent want realisation, the difference being that anti-activation strategies have to be employed whenever triggering conditions are in the offing, whereas want realisation can still be combated when the first-order want has already been activated. Unmediated strategies seem to have little chance here. On the contrary, attempting not to think of something you want tends to have the opposite effect.

3.4. Relations and Status of the Reflexive Structures

1. The four ways of relating optatively to oneself characteristically function in concert, that is, as various means to action or refraining from action. Not acquiring a want, managing to discard it or to prevent its triggering may have the same practical consequences as resisting the tendency to realise it. There can, however, be reasons for polarity divergences in a person's second-order wants of different orientations. Solomon's retention-oriented and activation-oriented wants are polarity-convergent, the latter only being required because of the (if he's lucky, short-term) inefficacy of the former. Emma, on the other hand, develops an activation-oriented want the polarity of which diverges from that of her retention-oriented want because, although she endorses her dispositional possession of the first-order empathetic attitude, she does not endorse its ubiquitous activation. She wants to keep it, but at the same time to keep it from running away with her. Acquisition-orientation and retention-orientation can likewise drift apart. Jock, for instance, deliberately developed the desire to go jogging regularly in order to keep himself healthy, but finds himself becoming dependent on running ever further and ever more often. As a result he wishes to be rid of this motivation. There is thus no conceptual reason why someone cannot become a slave to an autonomously acquired desire (cf. Mele 1993, pp. 271ff.; 1995, pp. 138ff.). Finally, as Frankfurt pointed out (1971, p. 16), acquisition-orientation and realisation-orientation can also be taken apart, as illustrated by his example of a drug therapist who wants to know what it is like to have the desires addicts have, without wanting to realise them.

2. A word is also in order about the status of mediating strategies. Their efficacy is a double-edged sword for a person's autonomy. Particularly where they are more than mere transitional instruments, they can take on the character of devices on which the person becomes dependent, thus restricting her self-governance. Someone who needs regular electro-shocks to prevent her from engaging in unwanted forms of behaviour is not governing herself relative to that behaviour, even if she goes in for it voluntarily and maybe even applies the shocks herself. But someone who manages to remain calm when a hated rival enters the room by quickly thinking of a peaceful seashore can plausibly be said to be governing herself by imaginative mediation. Unlike what might appear to be the case, the reason does not lie in the distinction between mental and physical strategies. Rather, the relevant difference lies in the extent to which the strategies thus employed obstruct the realisation of other of the person's wants, namely those somehow distinguished as particularly significant. If the agent's mediating strategy was merely to pin her self, that would hardly involve an impairment of self-governance, whereas if she needed to engage in hour-long meditation sessions every time her boss raised his voice, the mental character of the strategy would not qualify her behaviour as autonomous.

3. All in all, if the virtue of autonomy is a matter of being able to determine one's lower-order pro-attitudes in one of the four ways just sketched, then we are dealing with a characteristic that is both limited in its instantiation in empirical humans and subject to axiological relativisation in the light of other values. No matter how autonomous they are, humans are not able to construct their characters from scratch, reconstruct their characters in an optative vacuum, put indefinitely many of their dispositional wants on ice or resist any and every want that befalls them. Our autonomy is a limited and field-specific affair. And as a characteristic that is both reflexive and reflective, it is potentially in conflict with other features of human life that can also be subject to strong evaluation. Among these number passive phenomena, such as getting to know the contours of one's hedonic and optative dispositions and the joys of "letting oneself go", as well as non-reflective spontaneous activity.

4. Of-Authority and Valuing

1. In order to make sense of the idea of "self-rule", it has often seemed that we need criteria to distinguish those attitudes that belong to the "self" doing the ruling from those which don't. This is simply the search for the Of whose non-obstruction constitutes autonomy. There is widespread agreement (Thalberg 1989, p. 130; Watson 1975, p. 218; Frankfurt 1987, p. 166) that the mere fact that some want might be the highest in an optative hierarchy cannot install it at the heart of the agent's "self". This is, firstly, because a want of whatever order could always become the object of a yet higher-order want: a regress threatens. Second, even where this does not happen, there appears to be nothing about the embedding of one want inside another which could endow the non-embedded want with particular authority over the want thus embedded. If one want takes another as its object and opposes its retention, activation or realisation, then we have a semantic and attitudinal relation of subordination. But nothing guarantees that the subordinating attitude is the legitimate ruler over its subordinate and that, instead of autonomy, we are not dealing with a case of psychological repression or alienation.
2. A promising way of attempting to provide that guarantee involves picking out those higher-order wants that constitute the stance of valuing (Watson 1975, pp. 215f.; Bratman forthcoming). One reason for the plausibility of this suggestion is that, when we value something, we take on a stable attitude towards it. Moreover, although people's values change, a change in a person's values may be thought to involve a substantial change in the person herself. We thus have good reason to suppose that clarification of the concept of valuing is essential to an analysis of the virtue of autonomy. Further, should autonomy turn out to be analysable in terms of the realisation of one's values, then we would, so it seems, have discovered an excellent reason for investing autonomy itself with particularly strong value. Must we not inevitably value the realisation of our values, if we value anything at all?

4.1. Valuing

1. What, then, is it for a person to value something? In what follows I shall use the expression "A values x" as synonymous with "A cares for x", "x is important to A" or "x matters to A". I will sketch my suggestion with the help of another slightly revised version of empathetic Emma. Let us now imagine that she places a high value on helping others in need. What difference does this description make relative to the description of her as endorsing the retention of her first-order want to do so?

2. The first point to note is that the description of her as valuing helping others in need does indeed imply something pretty close, but not equivalent, to this. For her to value this proposition, it is necessary, first of all, that she have a first-order attitude, either a want or, as is more likely here, an emotion, which establishes a world-to-content direction of fit with the proposition that she help others in need.

3. Secondly, she cannot be said to value the object of her first-order attitude if she is indifferent to whether she continues to be concerned about that object (cf. Blackburn 1998, pp. 60ff.). Helping others in need would not be something she values if Emma were to be easily persuaded not to take her feelings about such people too seriously. However, that being the case does not require that she have reflected on her first-order attitude and explicitly endorsed retaining it. Of course, has she done so, that endorsement may well exclude such indifference, if the endorsement is itself backed by sufficient motivational force. But a reflexivity condition would be too strong a condition on valuing. People do not only begin to attach importance to things when their importance is questioned and consequently affirmed. A valuer must merely be disposed to give retention-oriented endorsement to her first-order attitude. That is, she must have at least a disposition to develop a second-order want — to be distinguished from already being the bearer of a dispositional (inactivated) want. A valuer would (if she hasn't already done so) form a relevant retention-oriented want under conditions that threaten her retention of the first-order concern, conditions which might very well remain counterfactual.10

4. Note that valuing requires the capacities both for reflective attention to one's desires and for the formation of higher-order wants directed at them. In order to attribute values and not mere pro-attitudes to an entity, we need to assume that the entity be able to distance itself both cognitively and optatively from its potentially motivating attitudes. Unlike autonomy however, valuing does not require that such reflection actually take place.

5. Finally, in order for us to be able legitimately to talk of valuing, the first-order attitude must, as I remarked, have some measure of stability. This stability may be either inherent in Emma's first-order disposition or else established by means of its second-order endorsement. Without such stability, a person's values could neither provide a standard for the criticism of her wants, nor would they display the feature of "a priori supervenience", according to which a person's valuing x necessarily involves her reacting consistently to properties of x that count as her reasons for so reacting.11

4.2. Valuing and Autonomy

1. According to the above analysis of valuing, there is a significant connection between valuing and one of the four forms of self-rule: valuing p entails either wanting to retain the want that p or being disposed to develop the want to retain the want that p.

2. However, although valuing, in the form described by the first disjunct, and retention-autonomy both involve the presence of a retention-oriented want, there is a causal difference between the two phenomena. Retention-autonomy requires that the retention-oriented want is causally effective. This is no implication of valuing, even where such a higher-order want is given. Emma is not autonomous relative to the possession of the desire to help others in need if she is unable to influence whether she continues to desire to do so. But if she endorses her continued desire to do so, then she values the giving of such assistance, independently of whether the rejection of that desire would lead her to discard it.

3. The closest connection between valuing and autonomy is given in cases in which people act both in accordance with what they value and because of valuing it, where doing so involves realising a want that is the object of retention-autonomy.

10 Note that, even where valuing does involve higher-order endorsement, what is valued is the object not of that second-order attitude, but of the ground-level want whose retention is thus endorsed. For this reason alone, valuing cannot, pace David Lewis (1989, p. 144), be identical to desiring to desire. As Harman (1993, pp. 150ff.) correctly points out, we don't value something towards which we have no factual pro-attitude. This is however no conclusive argument against the relevance of second-order wanting for valuing. In order to see this, we need clarity on the criterial significance, firstly, of retention-orientation and secondly, of the dispositional to form retention-oriented wants.

ent endorsement. If the higher-order endorsement is causally responsible for the existence of the lower-order want at the time of action, then the action is ceteris paribus — going to be autonomous.

4. The ceteris paribus conditions can however be lifted: it is perfectly possible that a first-order want someone stably and generally wants to retain can, under particular circumstances, interfere significantly with its bearer's autonomy. Although Emma wants to continue to desire to help others in need, that implies neither that she wants her empathetic disposition to be triggered every time she sees someone correctly thus describable nor that the attitude thus triggered inevitably have an effect on her action. What a person values concerns matters of importance to her on a fairly general level, i.e. under abstraction from specific contexts and the particulars of her involvement in them. And this leaves room in particular situations for a person's values to run away with her. The gap that can open up here grounds in the conceptual independence of activation- and realisation-orientation relative to retention-orientation. Where either of the former kinds of second-order rejection are given, retention-oriented endorsement is insufficient for autonomy.

5. For the same reason, valuing is also unnecessary for autonomy. People sometimes want to acquire, experience or realise desires whose objects they don't value in the slightest. An example of value-independent acquisition autonomy is Frankfurt's drug therapist, who develops the desire to acquire and experience the longing for an addictive drug. Value-independent realisation-autonomy is given when a (genuine) religious convert decides to follow his old sinful urges one last time. Note that there is nothing in the concept of higher-order wants that prevents them being singular in nature, that is, formed in and restricted in their application to one particular situation. Of course, the religious convert who forms such a singular want may find its scope spreading to take in many other situations that instantiates similar properties. Where this turns out to be the case, it might become correct to see him as caught between conflicting values. But singular cases, in which valuing and autonomy present opposing considerations, show that autonomy can involve no less than freedom from one's values.

6. Finally, perhaps the most obvious reason why the determination of our attitudes and actions by our values is insufficient for autonomy is that our valuing can itself be fundamentally non-autonomous. Leaving aside cases of value indoctrination, there are plenty of cases of valuing in which both the first-order concern and the disposition to desire its retention are firmly entrenched features of the person over which their bearer has no, or very little control. For most of us, that's the way we feel about avoiding being run over by motor vehicles. We disvalue being run down as firmly as we disvalue just about anything, but there's nothing autonomous about our taking steps to realise this value. That is, of course, no problem. We are quite happy not ruling ourselves here.

4.3. Autonomy and OF-Authority

1. Valuing, then, does not provide us with the OF whose non-obstruction constitutes the virtue of autonomy. Are we then left with a large hole in the middle of the concept? I don't think so. To see why not, let's return to the problems faced by a theory that identifies the autonomy-relevant OF with the highest-order attitude in a hierarchy of wants. The first problem noted at the beginning of §4 is the problem of a potential infinite regress. This seems to me of far less import than has been assumed. If the agent has reflected on her attitudes and has formed some higher-order attitude to which there are neither higher-order objections nor same-level competitors, then the mere possibility of an additional attitudinal level need not cause any problems. That is where the agent stands at that moment. Should she later reflect again and take on a new negative attitude of some kind to that want, then she will have restructured her optative life in such a way as to give a different feature OF-status.

2. The substantial problem may appear to be marked not by the regress objection, but by the authority objection. The authority of some specific attitude may be a temporally limited affair, as it may sooner or later be replaced by some other want. Why, though, should height of order be the arbiter of authority in the first place? Might not the higher-order want turn out to be the one that is in some sense alien to its bearer? Imagine now that Emma has been brought up not to act for "sentimental" reasons. That is, as a result of her upbringing, she is the bearer of a negative realisation-oriented want, which, although resistible, regularly leads her not to follow her impulse to help suffering creatures. Is it not counterintuitive to see her as autonomous when the higher-order principle leads her to repress her altruistic impulses?

3. Counterintuitive it may be, but one should avoid giving intuitions too much weight here that associate autonomy with moral virtue. Whether Emma is acting autonomously or would be better described as repressing a part of her self depends on the precise description of the case. If Emma never questions what she has been brought up to see as her duty, then she doesn't fulfil the condition of reflection required for autonomy. But if she does think these things through and has the ability to take a further effective optative stand on her second-order principle, but doesn't do so, then she satisfies the requirements and the principle acquires the status of an autonomy-relevant OF.

4. The requirement that the agent possess the ability to take a further effective optative stand on her highest-order want is essential. Without it, that highest-order want could itself be possessed heteronomously and thus be an unsuitable initiator of autonomy-conferring processes. Note, though, that a heteronomously possessed higher-order want need not be the product of a problematic genesis. On the contrary, it may simply result from reflection on some anthropologically basic first-order want, such as the desire for food or sleep (cf. §4.2.6.). However, such higher-order wants are likely to be excluded from generating autonomy anyway, as they are not generally going to bear the causal responsibility for their bearer's continued possession of the first-order attitude.

5. An accountable person is thus autonomous relative to a first-order pro-attitude where the following conditions are satisfied: she is the bearer of an optative con-
aptic; she has reflected on how she stands on that conflict; she has as a result formed a higher-order want of some specific orientation directed at one of the lower-level wants; that want is neither contested by other pro-attitudes on the same level nor rejected on a higher-level; and that non-rejection was itself accessible to change by a further optative stand in the agent's power. Where these conditions are fulfilled and the reflection satisfies minimal standards of rationality, the person is autonomous in relation to the relevant lower-order want. The presupposition of accountability and the use of an unanalysed concept of optative ability are two points at which the hierarchical analysis of autonomy is incomplete without a solution to the free will problem. But the responsibility for solving that problem was a burden that the hierarchical apparatus could never reasonably be expected to bear in the first place.

6. The virtue of autonomy both presupposes the basic form of freedom that makes us accountable and involves its exertion within a particular optative structure. Like any form of liberty, it opens up opportunities by creating obstructions to other factors. Autonomy involves a person setting up optative obstructions to her following her own blind impulses. Those obstructions are the effect of the constitution of an OF as a result of deliberation and the taking of a new optative stand. Having an autonomous character is having the inner freedom necessary in order to want or act in line with one's reflective optative attitudes, attitudes whose causal efficacy involves the overcoming of optative obstacles in the person herself.

7. At least in modern western culture, we tend to value the possession of the capacities to relate to oneself I have distinguished. Talk of an "autonomous character" plausibly becomes justified when the person is capable of exercising these reflexive capacities in a significant number of areas and with significant regularity. Autonomy is a (Human) virtue, in as far as such a character is a structural property whose retention we want, a retention-oriented want that we are in turn at least disposed to want to uphold. As a personal value, it no doubt stands in tension with moral values. Moreover, even within the sphere of personal values, it has a special status. This lies at least in part in the fact that it enables us, under certain circumstances, to act contrary to them—intentionally and without akrasia.

References


— (1987), "Identification and Wholeheartedness", in ibid., 159-176.


— (1785), Grundlage zur Metaphysik der Sitten, Prussian Academy Edition vol. 4, Berlin: Reimer 1911.


