

Penultimate version. June 2011.

Final version published In L. Radoilska (ed.), *Autonomy and Mental Disorder* (OUP, 2012), pp. 252–280

## **Autonomy and Ulysses Arrangements**

Lubomira Radoilska

In this chapter, I sketch the structure of a general concept of autonomy and then reply to possible objections with reference to Ulysses arrangements in psychiatry. In so doing, I commit to the claim that autonomy is best understood as a *de facto* contested concept rather than either an essentially contested one or an umbrella term covering a loose set of separate target concepts.<sup>1</sup>

The argument proceeds as follows. At the start of the inquiry, I explore three main strategies to conceiving autonomy in the current debate: value-neutral, value-laden, and relational. The objective is to bring into relief their distinctive rationales and draw attention to ensuing points of disagreement. Next, I identify two paradigm cases of autonomy by considering everyday instances of pre-commitment,<sup>2</sup> and then offer a first sketch of the concept of autonomy as opposed to the closely related freedom of action and intentional agency. Finally, I explain away the autonomy paradox to which the previously identified pair of paradigm cases seems to give

---

<sup>1</sup> Both the distinction between concepts and conceptions, and the notion of an essentially contested concept as employed in this chapter draw on Gallie (1956). On the thesis that autonomy is merely an umbrella term covering a loose set of separate target concepts, see Arpaly (2003, Ch. 4). The differences between the two rejected models of understanding autonomy – an essentially contested concept and an umbrella term – are of no bearing to the present inquiry, since both models have similarly sceptical implications with respect to the project of articulating a general concept of autonomy. By sketching such a concept, it becomes clear that neither of these models is persuasive. On the importance of locating competing autonomy accounts with respect to a possible general concept, see Ch. 1 of this volume by Jane Heal.

<sup>2</sup> Unless stated otherwise (see in particular the penultimate section), the terms ‘pre-commitment’ and ‘Ulysses arrangement’ will be used interchangeably throughout the chapter.

rise in the context of mental disorder. By resolving this paradox, we learn two valuable lessons. The first is about the relationships between the three conceptions of autonomy mentioned above. The second is about the relationships between autonomy and mental disorder.

### **Three approaches to autonomy**

A plausible way to map out the broader lines of disagreement on autonomy is to compare three main alternatives that have emerged from the current debate. These alternatives are often referred to as value-neutral, value-laden, and relational strategies (Christman 2003; Taylor 2005; Buss 2008).<sup>3</sup> The present section will look into each of these strategies in turn.

The first, value-neutral approach focuses on identifying the formal features that distinguish autonomous from non-autonomous actions and choices. These features are meant to capture the special relationship that a person has to a particular category of her actions and motives that could be seen as her own in a suitably robust sense. This point becomes clear, if we consider cases where the autonomy conferring relationship has broken down, such as instances of coercion and compulsion. A person subjected to coercion does not act by choice. On this occasion, her motives to act are not fully her own. Similarly, a person in the grips of compulsion does not act by choice. In this respect, her motives are not sufficiently up to her.<sup>4</sup>

These two cases illustrate the kinds of constraints that need to be eliminated in order for a choice to be considered as autonomous: external constraints endangering a person's ownership over her

---

<sup>3</sup> The terms 'value-neutral' and 'procedural' are sometimes used interchangeably with respect to autonomy conceptions. The same goes for 'value-laden' and 'substantive'.

<sup>4</sup> Compare Robert Nozick's analysis of coercion as a successful attempt to modify a person's intentions by means of credible threats (1969) and Lennart Nordenfelt's account of compulsion as subjective unavoidability (2007).

motives and internal constraints belying her authorship of these motives.<sup>5</sup> Value-neutral accounts define the relationship to one's motives that is free from these and similar constraints in a variety of ways.<sup>6</sup> According to a classical approach, the underlying attitude is captured in terms of endorsement, identification or, more modestly, acceptance (Frankfurt 1971; 1998, Ch. 12; 1999, Ch. 11; and 2002). An alternative approach points to the significance of a historical perspective onto the formation of both first-order motives and higher-order attitudes, such as endorsement (Christman 2007). This further criterion aims to address less apparent threats to autonomy that may not be recognised as such by the person affected. Examples include so-called adaptive preferences, where the desirability of different options is inappropriately influenced by their perceived (un)availability as in Aesop's fable 'The Fox and the Grapes' (Elster 1983; Colburn 2011).

These variations notwithstanding, both alternatives concur on a central idea, according to which the autonomy conferring relationship of a person to her motives can be fully defined with no reference to further principles or values. A major implication is that autonomy is presented as a separate source of normativity, on a par with moral and prudential considerations, such as beneficence and efficiency.<sup>7</sup> The underlying reasoning can be reconstructed as follows:

---

<sup>5</sup> See Feinberg (1980, Ch. 1; and 1986, Ch. 18) and Moran (2001). See also Ch. 3 where K.W.M. Fulford and I discuss the Aristotelian notion of being 'up to the agent' which underlies considerations of both ownership and authorship with respect to motives and actions.

<sup>6</sup> In the following, I shall comment on two alternatives only, for the aim is to pick out the underlying strategy that value-neutral accounts share rather than detail possible points of disagreement within it.

<sup>7</sup> On the notion of sources of normativity, see Korsgaard (1996). The interpretation of value-neutral autonomy as one source of normativity among others that I propose departs from Christine Korsgaard's thesis, according to which autonomy is the ultimate source of normativity, since it underwrites a person's ability to put herself under obligation and to perceive certain

- 1) Autonomous choices deserve special consideration simply by virtue of being a person's own in a suitably robust sense, such as being free from undue influences, both external and internal.
- 2) The significance of autonomous choices as defined above cannot be duly acknowledged unless there is a protected sphere of action, where these kinds of choices may be followed through without further justification.
- 3) Within this sphere of action, an autonomous choice should be treated as authoritative, although it may happen to be morally objectionable or unwise.

The intuitive appeal of this line of thought is reflected in the liberal idea of toleration as a prerequisite for personal freedom within society. The implicit links between autonomy and toleration become clear once we notice that respect for either comes to matter in similar circumstances: whenever a particular choice faces solid objections and these can be enforced over dissenters. As Joseph Raz points out, we cannot tolerate other people's virtue but only their limitations (1987, p. 320). Similarly, respect for autonomy would be superfluous with regard to choices that are generally considered as admirable. In this regard, autonomy and toleration offer two complementary ways to define one and the same cluster of choices as particularly exposed to interference and worth protecting. The first point stems from the fact that the choices at issue are deemed to be objectionable in some conspicuous way, such as being foolish, perverse, or wrong, to paraphrase John Stuart Mill (1859, p. 53). An intuitive reaction would be to try and undo these kinds of choices. The second point expresses a core insight of the value-neutral approach indicating that there must be a sphere of action where autonomous choices need not be backed up by further reasons. It requires that the initial reaction in favour of interference is curtailed, but

---

actions as worth undertaking. This point will be further developed in the subsequent discussion on value-laden accounts of autonomy.

does not remove the grounds for disapproval that motivated it in the first instance. This is why toleration may appear paradoxical at first sight. For it involves both a strong negative evaluation and a commitment not to act upon it (Scanlon 1996). The idea of autonomy as an independent source of normativity helps dissipate the impression that toleration is odd. The underlying intuition is clearly expressed in Mill's classical treatise *On Liberty*:

“If a person possesses any tolerable amount of common sense and experience, his own mode of laying out his existence is the best, not because it is the best in itself, but because it is his own mode.” (1859, p. 114)

Turning to the second key approach to autonomy, we are in a position to notice an immediate contrast. For value-laden accounts intertwine autonomy and practical rationality and argue that responsiveness to reasons is the distinctive feature of autonomous choices (Hill 1991; Korsgaard 1996; Watson 1975; 2004). A direct implication is the idea that, ultimately, the source of normativity can only be one. It is supported by an intuitive line of reasoning that could be summed up as follows:

- 1) Autonomy could hardly be a plausible normative concept if its task were to uphold morally objectionable choices or prop up other forms of practical irrationality;
- 2) Instead, choices are worthy of special consideration only if they respond to reasons, the validity of which could also be appreciated from an impartial or ideal observer's viewpoint;
- 3) A choice is autonomous to the extent that it is responsive to reasons in the sense specified above rather than to mere incentives.<sup>8</sup>

The value-laden interpretation of autonomy as practical rationality and ultimate source of normativity could be grounded in a strong analogy between motivation and cognition. This

---

<sup>8</sup> The contrast between reasons and incentives draws on Kant's distinction between motives and incentives (*Groundwork* 4: 427).

analogy is well captured by the concept of ‘orthonomy’ or correct determination as applicable to desires (Pettit and Smith 1996). To bring out this point, let us consider, following Pettit and Smith, the nature and scope of freedom in theoretical settings. For instance, our intellectual freedom would be effectively undermined by a licence to brave the laws of logic, believe in patent falsehoods, and ignore evidence on hand. These examples suggest that, in the realm of belief, freedom presupposes rules and constraints that forestall hindrances to reliably getting things right, such as wishful thinking and hasty conclusions. In fact, following these rules and respecting these constraints are key expressions of our freedom of thought. Conversely, forming a belief without sufficient reason and refusing to revise an earlier opinion in the face of contrary evidence are paradigm cases of having one’s freedom of thought compromised. The core idea is that what holds her back as a thinker boils down to an inadequate relationship between beliefs and supporting reasons. From this perspective a belief is unfree to the extent that it is unwarranted.

Drawing on this account of freedom with respect to beliefs, it is plausible to conclude that freedom with respect to desires would be equally undermined by a licence to satisfy any fancy, ignore reasonable demands, or abandon long-term commitments on a whim. For the kind of volitional arbitrariness these examples have in common structurally resembles the inadequate relationship between beliefs and supporting reasons that we have seen to erode freedom of thought. Following this line of reasoning, it is natural to stipulate that freedom of will should depend in a similar way on a fitting relationship between motives and reasons for action. This relationship would only obtain if motives are responsive to reasons as opposed to what we may call unexamined incentives.

The notion of orthonomy offers an insightful way of thinking about freedom with respect to desires. In particular, it enables us to identify anything that holds a person back from engaging in worthwhile pursuits as an obstacle to her freedom. On this view, a choice is unfree to the extent that it is idiosyncratic rather than responsive to reasons. The outcome leaves no room for a

principled asymmetry between first- and other-person perspectives on individual choices. In doing so, it corroborates the underlying assumption of value-laden accounts of autonomy, which is to reject the possibility of alternative sources of normativity.

Orthonomy is not the only possible interpretation of responsiveness to reasons. This essential idea could be also presented in Kantian terms, such as universalizability of our motives for action or ability to recognise and fulfil unconditional requirements (Hill 1991; Korsgaard 1996). In each instance, however, the main ambition is relevantly similar, that is, to contrast autonomy with various forms of arbitrary, idiosyncratic, or otherwise indefensible expressions of the will, independently of whether they have been initiated by the agent herself or prompted by a third party.

The third major strategy to defining autonomy offers a critical standpoint to both the preceding value-laden approach and its value-neutral rivals. Its essential idea is that interpersonal relationships are constitutive of personal autonomy rather than contingent factors that may impede or facilitate it (Meyers 1989; Stoljar 2000). This leads to a distinctive claim, according to which personal autonomy is best presented in terms of a particular social-relational status. As Marina Oshana puts it:

“... to be autonomous is to stand in a certain position of authority over one’s life with respect to others” (2006, p. 94).

In support of this claim, relational accounts argue that unless we take on an interpersonal perspective on autonomy, we are bound to overlook the effects of oppressive arrangements on an oppressed person’s capacity for autonomy. As a result, we would end up with a flawed distinction between autonomous and non-autonomous choices that effectively condones the loss of personal autonomy incurred by members of dominated social groups.

A related worry is that socialisation into a deferential role typically leads to earnest acceptance so that a person thus socialised naturally gives way to the preferences of powerful others and, moreover, does so without a trace of resentment. This is not to say that there is something inherently suspect in motivations that depart from self-interest. Nor is the idea to suggest that personal ambition is a prerequisite for personal autonomy. Instead, the point is to draw attention to a pervasive way in which imposed choices could be mistaken for autonomous from both first- and other-person perspectives.

This conclusion could be disputed. For both value-neutral and value-laden accounts seem in a position to address the psychological threats that oppressive socialisation poses to autonomy. For instance, a value-neutral notion of adaptive preferences would be able to identify as non-autonomous individual choices traceable to a systematically abusive environment. A value-laden approach to autonomy could also spot that autonomy is being compromised in such instances because the choices at issue are clearly not responsive to reasons, but manipulated from outside.

Although persuasive, this line of reasoning does not tackle the main challenge raised by the relational strategy, which is to account for the social nature of autonomy rather than concentrate on psychological manifestations that may turn out to be secondary. The thought is that what goes wrong in circumstances of oppression has to do primarily with the social-relational status of the person oppressed rather than the way she relates to her motives or is able to track valid reasons for action. In other words, an oppressed person lacks autonomy by virtue of being oppressed, independently of how candid she is about her reasons for action and whether, counterfactually, she would have chosen to do the very same things that are now imposed on her. If this is correct, a focus on the psychology rather than politics of autonomy would obscure the real threats to autonomy. Moreover, it would make them appear as personal rather than social problems. This result is paradoxical in so far as it implies that an oppressed person should take it upon herself to become autonomous whilst being oppressed in the same way as before. The relational strategy to



autonomy aims to resolve this paradox by adopting a social and interpersonal stance in opposition to both the intrapersonal and, in a sense, impersonal viewpoint championed respectively by the value-neutral and value-laden approaches.

### **Ulysses arrangements as paradigm cases of self-determination**

Having set out these three approaches in general terms, let us consider their possible interactions in the context of a challenging case where competing concerns about autonomy appear to be equally plausible. This case relates to the use of Ulysses arrangements in psychiatry. But before fleshing out the autonomy paradox, to which this use seems to give rise, I shall look in some detail into paradigm cases of pre-commitment outside psychiatry. The idea is to establish whether there is something paradoxical in the very notion of pre-commitment, in advance of discussing its possible applications to contexts of mental disorder.

Drawing on works by Ion Elster (1984; 2000) and Michael Bratman (1999; 2007), it is persuasive to interpret pre-commitment broadly conceived as an essential aspect of intentional agency over time. This becomes clear as soon as we think of everyday planning, which requires routine decision-making about future and counterfactual situations just as much as immediate action. For instance, planning to go to the cinema next weekend requires not only that I book my ticket when I make up my mind, say, on Tuesday (immediate action), but also that a few days later I try to turn up for the screening I booked (commitment to a particular course of action in the future). A final element of my planning is to rule out foreseeable alternatives, such as staying at home or taking up last minute offers for a weekend break (commitment to stick to the plan). This latter commitment not only gives my plan to go to the cinema a better chance to succeed, but effectively distinguishes it as a plan of mine rather than a mere fall-back option that I create for myself, i.e. booking a cinema ticket for the weekend on Tuesday in case I have nothing better to do over the weekend. In contrast, booking a cinema ticket for the weekend on Tuesday amounts

to planning in so far as I pre-commit to both try to attend the screening I booked and resist obvious alternatives, like staying at home. With respect to my planned cinema going, such alternatives count as ‘distractions’ or even ‘temptations’. Opting for them equals failure for my cinema project.

In fact, it is in such circumstances – various failures or at least difficulties to stick to the plan – that the role of pre-commitment for intentional agency over time becomes salient. To return to the example above, my booking of a cinema ticket for a screening a few days later usually suffices to keep in check ‘distractions’, such as arbitrary changes of mind and sheer procrastination when it is time to get ready and go out. Moreover, my pre-commitment to resist such distractions and go to the cinema instead may not even come to attention as long as I go ahead with the plan wholeheartedly and am not the least attracted by alternative courses of action. Conversely, I may book my ticket in advance not because I am swayed by the idea of going to the pictures, but in order to make sure that, at the weekend, I will have good reason to go out rather than stay at home as usual. In this second scenario, pre-commitment comes to the forefront of my thinking because I wish to bring my plan to fruition, whilst being aware that my commitment to it is less than wholehearted. By booking a cinema ticket in advance, I give myself additional leverage against a temptation I foresee. As a result, I am able to realise a plan that I am otherwise unlikely to follow through. In this respect, pre-commitment broadens the scope of my freedom to act and opens up avenues that were previously closed or nearly closed for me.

This function of pre-commitment is well illustrated by Ulysses’ encounter with the Sirens which, following Elster (1984), I propose to consider as a paradigm case. Drawing on the *Odyssey*, the options available to Ulysses without pre-commitment are two: either to sail away from the Isle of the Sirens and not hear their enchanting song, or to approach them and perish, for this song lures the sailors to their deaths. However, by implementing an ingenious form of pre-commitment (Ulysses orders his crew to tie him to the mast and plugs their ears up with beeswax), he is able to

create a new option: to hear the song and not be lured to his death. Let us call this function of pre-commitment trailblazing.

There is a further function of pre-commitment relevant to the present inquiry that may slip under the radar if the focus stays with individual actions rather than agency over time. In this respect, the second cinema going scenario offers a better reference point than Ulysses and the Sirens. This is because it brings into relief a separate way in which resilience against what I see as temptation may increase my freedom. Here the benefit of pre-commitment is not so much that it expands the range of options from which to choose, but that it enables me to exert indirect influence over the kind of person I will be in the future. As the reluctant cinema going example indicates, this can be achieved by devising courses of action (booking my cinema ticket in advance) to counteract emerging trends in my behaviour, which I dislike (doing nothing over the weekends). In this case, pre-commitment is undertaken not as a one-off, but as a series of actions that I plan in advance in such a way as to make difficult to abandon for the sake of foreseeable and strong temptations. To return to the previous example, if I am worried about becoming inactive as soon as my working hours are over, booking a cinema ticket for the weekend on Tuesday is part of a bigger project, finding things to do over the weekends. If this kind of pre-commitment is successful, it will help to gradually change my routine to the point that no further pre-commitment will be necessary in order for me to engage in various exciting activities in my spare time. This is because, with respect to leisure, I would have become the kind of person I wanted to be when I began reshaping my weekend habits by ensuring that, albeit reluctantly, I would go to the cinema for the sake of a prepaid ticket.

This second function of pre-commitment is well accounted for by Aristotle's theory of ethical habituation as outlined in book II of *Nicomachean Ethics*.<sup>9</sup> Drawing on this account, a person is able to develop virtuous character traits by engaging into praiseworthy activities in spite of the fact that at the beginning they have no immediate attraction to her. This is particularly salient if we think of acts of courage and temperance as they both require the suppression of strong intuitive responses, to avoid imminent danger in the former case and to forego instant gratification in the latter. The ability to do so may seem unsurprising once the relevant virtues, e.g. courage and temperance have been developed. For, a person of virtue considers virtuous actions not only as admirable but also as agreeable to perform. The question is how anyone could perform such actions without already being of virtuous character. Pre-commitment offers a plausible answer. Thus, although I am bound to perceive acting in accordance with virtue as onerous at the start of my ethical habituation, the fact that I live in a human society puts me in a good position to take control over my intuitive reactions to apparent danger and promised pleasures. In particular, the anticipated disapproval of others, especially significant ones, offers me – under the guise of a sense of shame – the required motivational leverage so that I make myself more resilient to distractions that could induce me to act in a cowardly or an intemperate manner. If my ethical habituation is successful, I grow out of my sense of shame and begin to act in accordance with virtue for virtue's sake. Disapproval avoidance no longer figures among my reasons for doing the right thing. In fact, it is natural to expect that in so far as I am a person of virtue, I will do so even in the face of strong though misguided societal disapproval.

The preceding considerations point to a second paradigm case where the lead function of pre-commitment is no longer trailblazing but what I propose to call character-building. The focus here is on moulding the kind of person I am turning into over time rather than enabling the

---

<sup>9</sup> I have developed the subsequent point in more detail elsewhere, see Radoilska (2007, pp. 233–290).

performance of a particular kind of action in the future. Clearly, trailblazing and character-building are not mutually exclusive. Quite the reverse, many instances of pre-commitment would combine both as in the case of booking a cinema ticket whereby I not only give myself an extra reason to go to the pictures this weekend (pre-commitment as trailblazing) but also start getting myself into a more proactive mind-set (pre-commitment as character-building).

Having identified and explored these two paradigm cases of pre-commitment, we are in a position to see that neither of them is likely to lead to an autonomy paradox in the sense specified at the start of the section, that is, to contribute to autonomy on one plausible conception, but be hostile to it on another, and beside the point on a third. This is because the two central cases of pre-commitment, trailblazing and character-building are at the same time central cases of active self-determination. In both instances, a person steps in and takes charge of her situation instead of accepting it as a given. Thus, by means of trailblazing, I make possible the course of action I wish to undertake rather than go along with the range of pre-existing options. In so doing, I expand my domain of self-determination. Similarly, by means of character-building, I work my way up to the kind of person I would like to be rather than settle for the character I inadvertently turn into, e.g. a ‘couch potato’, to return to the reluctant cinema going example. In so doing, I strengthen my capacity for self-determination.

Moreover, by considering these two central cases of pre-commitment, we are able to observe a particular feature that distinguishes active self-determination from two closely related categories, freedom of action and intentional agency. Both are consistent with taking up an option from a pre-set range about which one has no say, a mode of self-determination I propose to call passive. Furthermore, if the course of action I would like to undertake is within the range of options available to me and if I am happy with the kind of person I am becoming, there is no reason for me to engage in active self-determination as a separate project, in addition to what I already do. In fact, to the extent that they meet with success, the two discernible forms of active self-

determination, trailblazing and character-building, are also bound to slip back into the background of intentional agency. As suggested by the preceding discussion of pre-commitment, active self-determination becomes salient only when things risk getting out of hand.<sup>10</sup>

So there are at least three categories of self-determination to be distinguished: firstly, passive self-determination, which is consistent with freedom of action and intentional agency but not autonomy; secondly, inconspicuous active self-determination or trouble-free autonomy; and thirdly, active self-determination in the spotlight or autonomy as express pre-commitment, which takes the form of trailblazing, character-building, or both.

Given the centrality of express pre-commitment, it is plausible to consider aptitude to account for its two main forms as a test for eligible autonomy conceptions. More precisely, if a conception treats either of these forms as irrelevant or hostile to autonomy, this would count as a strong reason against this conception. Let us now turn to the three autonomy conceptions I presented in the previous section: value-neutral, value-laden, and relational, and see how each of them fares on this test.

Firstly, a value-neutral conception would have no difficulty accounting for either paradigm case since it already considers autonomy in terms of a dual relationship of ownership and authorship applicable to both actions and motives. As clarified in the previous section, this is to reflect that possible constraints to self-determination may be external, that is, ownership affecting or internal, that is, authorship affecting. Trailblazing and character-building enable a person to address both types of constraints, the first at the level of actions and the second at the level of motives. For instance, trailblazing provides us with means to act only on motives that we endorse and to resist desires that we may find puzzling, disturbing, or repugnant. This type of pre-commitment is

---

<sup>10</sup> For a related point, see Chapter 9 by Hallvard Lillehammer.

consistent with a Frankfurtian notion of identification as the hallmark of autonomous actions.<sup>11</sup> Similarly, character-building enables us to affect our underlying motivations rather than relate to them as ‘passive bystanders’ (cf. Frankfurt 1971). By allowing us to remove various obstacles to becoming a well-integrated self, such as ambivalence, indifference, and other forms of self-alienation, this type of pre-commitment concurs with Frankfurt’s analysis of ‘wholeheartedness’ as a distinctive feature of autonomous persons (1998, Ch. 12).

Secondly, a value-laden conception would be able to accommodate the two paradigm forms of pre-commitment to the extent that they enable a person to respond to good reasons as opposed to mere incentives. For instance, trailblazing would amount to exercise of autonomy only if the option I aim to resist is bad or unworthy in some objective sense, over and above its being incompatible with a plan that matters to me. To return to the cinema example, my advance booking of a ticket so that I do not stay at home over the weekend would be a plausible exercise of self-determination only if it is bad to spend one’s weekends this way, not just something I happen to dislike.<sup>12</sup> In a similar vein, character-building would count as autonomous only if the dispositions or traits I aim to develop are good, in addition to being aspirational.

Finally, a relational conception would be in a position to appreciate trailblazing and character-building to the extent that they both put the spotlight on interpersonal dynamics as key aspect of self-determination. In this respect, the cinema example would not be considered of relevance to autonomy, at least not without providing further details about the social-relational background of the person who engages in pre-commitment. However, both the original Ulysses arrangement

---

<sup>11</sup> See, in particular, Frankfurt (1998, Ch. 5).

<sup>12</sup> Watson (2004, Ch.3) offers an example of this line of value-laden reasoning, since a central claim is that addiction is incompatible with autonomy not because it typically jeopardises projects that a person with addiction cares about but because it allows ‘disordered appetites’ and unworthy commitments to become the centre of her life.

(trailblazing) and the Aristotelian account of moral development I sketched earlier (character-building) could be adopted as central instances of relational autonomy, since they aptly point to the significance of constructive and trusting interpersonal situations. Thus, Ulysses would have been unable to safely hear the Sirens' song if his companions could not be trusted to both tie him to the mast as arranged and release him afterwards. Similarly, a person's moral development would be severely hampered if the community in which she grows up cannot be relied upon for ethical guidance. Both points are in line with the core insight of relational approaches to autonomy highlighting that oppressive social relationships make self-determination impossible, not merely difficult, whenever they prevent us from developing empowering kinds of reliance on others, like the original Ulysses arrangement and an Aristotelian character-building (cf. Stoljar 2000).

So, all three types of autonomy conceptions are consistent with the centrality of express pre-commitment and could therefore be considered as conceptions of the general concept I began to sketch in this section, that is, autonomy as active self-determination. This is a promising ground for both addressing apparent incongruities between different autonomy conceptions and distinguishing them from conceptions of neighbouring concepts, such as freedom of action and intentional agency.

However, to pursue this line of inquiry we should first resolve a difficulty raised by some discussions of pre-commitment in psychiatric contexts. The worry is that these discussions support the conclusion that express pre-commitment is an autonomy paradox in the sense ruled out by the preceding analysis. If correct, this conclusion would leave us with a dilemma: either to reject the general autonomy concept outlined or to reject some of its conceptions we just acknowledged. Either option would undermine the argument proposed here.

### **Ulysses arrangements as a possible paradox of self-determination**



The challenge anticipated at the end of the previous section seems particularly to the point since the instances of pre-commitment in psychiatry that give rise to the putative autonomy paradox match the central cases of active self-determination, trailblazing and character-building I described earlier.

The first instance relates to advance directives anticipating various situations that could arise in later stages of dementia when a person is no longer able to make or express competent decisions.

<sup>13</sup> It is compelling to consider this kind of pre-commitment as an obvious case of trailblazing, for it extends this person's range of options to encompass choices which would otherwise be outside her control.

The second instance has to do with the use of Ulysses arrangements in order to anticipate various situations, in which a person loses insight into the effects of mental disorder on her thinking, but still fulfils the criteria for decisional capacity.<sup>14</sup> As a result, she is likely to make legally competent yet puzzling decisions, which she cannot even recognise as her own on recovery. Key Redfield Jamison's memoir of an expert on bipolar disorder who is also living with this condition offers a poignant testimony to this effect (1995). Drawing on her account, Ulysses arrangements offer a particularly efficient way to forestall related threats to self-determination, such as acting out of character during manic episodes. This is because these arrangements enable a person to appoint trusted others, including professionals and personal relations so that they may act as her guardians for the duration of such periods of confusion. The direct link with the second paradigm case of pre-commitment, character-building becomes apparent as soon as we take into

---

<sup>13</sup> Consider, for instance, the situation of Mrs Day, a person with dementia who lacks decisional capacity and whose care has to be agreed by third parties, including healthcare professionals and relatives because she has not drafted advance directives whilst still having capacity (Ch. 8 by Elizabeth Fistein). On the notion of decisional capacity, see the Introduction to this volume.

<sup>14</sup> On the relationship between capacity and insight, see Ch. 7 by Jules Holroyd.

consideration the tendency of this kind of Ulysses arrangements not merely to pre-empt the implementation of perplexing decisions (immediate objective), but also to remove the need for pre-emption (indirect objective). To return to the memoir above, having specified the conditions under which she could be given involuntary psychiatric treatment in a Ulysses arrangement, Jamison was able to spot the warning signs for herself so that it never had to be enforced in the following two decades. This is confirmed by recent research on the psychiatric treatment received by people with bipolar disorder who have previously signed similar kinds of Ulysses arrangements (Gremmen 2008), which reported a steady, statistically significant correlation with improved insight and almost exclusively voluntary treatment in contrast to people with the same condition who have not signed such an arrangement. Both Jamison's experience and this further evidence are consistent with the analysis of successful character-building I proposed in the previous section.

In light of these observations, it is safe to conclude that there is no relevant difference between, on the one hand, these two uses of pre-commitment in psychiatry and, on the other, the two paradigm cases of active self-determination in general. This is why the thought that these kinds of pre-commitment may lead to an autonomy paradox in psychiatry deserves careful consideration.

With respect to the first kind of pre-commitment, advance directives re later stages of dementia, this thought follows from a recent critique of the practice put forth by Agnieszka Jaworska (1999). According to Jaworska, the advance directives of people with dementia who no longer have decisional capacity should not be taken for an ultimate expression of their self-determination but may be overridden by carers and family to protect the interests that these people appear to have developed in their new situation. This central claim rests on a broader account of autonomous agency, which ties the latter in with a capacity to value or deeply care about things as opposed to a capacity to implement one's values, which is deemed secondary. Drawing on empirical evidence, Jaworska suggests that, although people with dementia eventually lose the

latter capacity, which she considers as inessential for autonomy, they still retain the former, which, on her account, is autonomy-conferring. So advance directives re later stages of dementia lead to an autonomy paradox on this view. This paradox is as follows: if this kind of pre-commitment is honoured, the people whose self-determination it means to uphold are denied self-determination. This is because, instead of being supported to implement the projects they have come to care about, people in the later stages of dementia are made to live by the decisions of their earlier, more articulate selves with whom they no longer identify.

The interest of this challenge becomes apparent as soon as its implicit reliance on a relational approach to autonomy comes to attention. For the underlying intuitions that motivate Jaworska's critique seem to be the following. Firstly, whenever respect for a person's self-determination is confined to sticking to her earlier instructions as opposed to giving credence to her current preferences, she is unavoidably attributed lesser standing than that of a fully autonomous person who can revisit her prior commitments at any point.<sup>15</sup> Secondly, the assumption of lesser autonomy is self-fulfilling for it leads to closing down opportunities for self-determination that this person could have otherwise taken up.

With respect to the second kind of pre-commitment, Ulysses arrangements re manic episodes, the conclusion that they lead to an autonomy paradox is reached in a recent discussion by Theo van Willigenburg and Patrick Delaere (2005). Unlike the preceding critique, the underlying concern here is not so much that Ulysses arrangements are inimical to autonomy, but that they are beside the point. The supporting argument builds on the thesis that the kind of self-control that these arrangements promote does not amount to exercise of self-determination. This is because this kind of pre-commitment allows only for self-control in terms of self-enforcement, whereas self-determination requires a different category of self-control, that is, self-legislation. Although in

---

<sup>15</sup> For a compelling analysis of changing one's mind as an inherent aspect of self-determination, see Radden (2005).

both instances the goal is set by the person herself as opposed to an interfering third party, the authors reason, there is a significant difference between the ways, in which she goes about realising this goal. In the case of self-legislation, she readily carries out her plan, whilst in the case of self-enforcement she merely goes along with it, since it has become impossible for her to pull out. Following this line of reasoning, Ulysses arrangements could be seen as paradoxical. For they turn out to be irrelevant to what the authors consider as autonomy in the strict sense (self-legislation) and at the same time preserving what they call autonomy as authenticity (self-enforcement).

The interest of this second challenge could be better appreciated if we bring into relief the value-laden approach to autonomy, on which it apparently rests. More precisely, Kant's distinction between acting from duty and acting in conformity with duty<sup>16</sup> seems to yield support to a stark contrast between self-legislation and self-enforcement. The point could be that, as a kind of self-enforcement, Ulysses arrangements are consistent only with acting in conformity with duty but not with acting from duty. Yet, only the latter mode of action is (truly) autonomous, for it responds directly to reasons as opposed to mere incentives. So the autonomy paradox, to which Ulysses arrangements give rise on van Willingeburg and Delaere's view, has to do with the supposition that self-enforcement does not leave room for responsiveness to reasons since it becomes relevant when only incentives, but no good reasons for action impress on an agent.

To recap, the conjunction of these two critiques of pre-commitment in psychiatry leaves us with the following paradox: instead of being central cases of active self-determination where alternative approaches to autonomy converge, trailblazing (in the guise of advance directives re dementia) and character-building (in the guise of Ulysses arrangements re manic episodes) look like essentially contested cases displaying how unbridgeable the gaps between alternatives

---

<sup>16</sup> *Groundwork* 4: 398–9. See also: Timmermann (2009).

effectively are. Arguably, the first is to be dismissed as compromising self-determination proper on relational grounds, whereas the second seems to be beside the point from a value-laden perspective. It looks as though only if a value-neutral view is adopted, both cases remain steadily associated with self-determination.<sup>17</sup>

My reply to this challenge will take two steps. The first will be to explain away the paradox above by identifying the mistake on which it rests. The second will be to expand on the valuable lessons about autonomy and its relationship to mental disorder that we learn from this mistake.

The mistake in question becomes apparent as soon as we recall a distinction between two ways of exercising autonomy, which emerged from the analysis of pre-commitment in the previous section: on the one hand, trouble-free autonomy, where the workings of pre-commitment remain inconspicuous, and, on the other, autonomy as express pre-commitment in view of considerable threats to the authorship or ownership of one's actions or motives. As argued earlier, both ways of self-determination share the same structure, and the benefit of explicit forms of pre-commitment is to help us see clearly this structure which is otherwise barely distinguishable from the related intentional agency and freedom of action.

With the distinction between trouble-free autonomy and autonomy as express pre-commitment in mind, let us now return to the disputed conclusion that pre-commitment in psychiatry is a paradoxical form of self-determination. Clearly, this conclusion hangs on the idea that something akin to trouble-free autonomy is the genuine item, with respect to which express pre-commitment

---

<sup>17</sup> See, for instance, Ronald Dworkin's defence of advance directives and the underlying 'integrity view of autonomy', according to which the right to autonomy is meant to protect "the ability to act out of genuine preference or character or conviction or a sense of self" (1993, p. 225). I have not expanded on this or alternative value-neutral discussions of pre-commitment in psychiatry since they are fully consistent with my analysis of pre-commitment as paradigm case of self-determination and, therefore, are not conducive to the paradox under consideration.

falls short. This is because, in order to make sense of the concern that pre-commitment in psychiatry may promote some lesser kind of self-determination at the expense of what is perceived as self-determination proper, a concern expressed by both Jaworska (1999) and van Willigenburg and Delaere (2005), we need to conceptualise express pre-commitment as an imperfect and potentially misleading replica of trouble-free autonomy.

However, there is good reason to reject this conceptualisation, namely that it effectively misplaces the locus of autonomy from active self-determination to effortlessness. The upshot is unavoidable once trouble-free autonomy is set out as a standard. Express pre-commitment is bound to fail this standard since it is a trouble-shooting exercise of autonomy. The problem is not that, following this conceptualisation, we end up with a view of autonomy that is unduly restrictive but that this view is simply mistaken. By focusing on effortlessness, which distinguishes trouble-free autonomy from autonomy as express pre-commitment, we not only begin to lose sight of the underlying feature they share, namely, that they are both instances of active self-determination. Moreover, the distinction between trouble-free autonomy and passive self-determination becomes blurred, for effortlessness is also typical of the latter, yet it is not compatible with autonomy. In fact, building on the point I made earlier about passive self-determination, we could say that its effortlessness boils down to making no effort toward self-determination. In contrast, the effortlessness of trouble-free autonomy is the outcome of successful self-determination. The apparent absence of obstacles to self-determination in this latter case can be traced back to a series of effective pre-commitments whereby a person has managed to influence the range of options open to her (trailblazing) as well as develop stable dispositions she approves of (character-building).<sup>18</sup> By looking in isolation at instances where no

---

<sup>18</sup> Aristotle's account of eudaimonia or an accomplished life of virtue offers a good illustration of trouble-free autonomy. See, in particular, *Nicomachean Ethics* 9.4 which expands on virtue as a state of friendship with oneself. A distinctive feature of this state is that there is nothing a person

obstacles to self-determination seem to be present, it is easy to take the effortlessness of active self-determination that has become trouble-free, that is, the effortlessness of an achievement for the effortlessness of passive self-determination, that is, the effortlessness of not giving it a go.

The upshot of shifting attention away from active self-determination toward effortlessness can be observed in two related moves that critics of pre-commitment in psychiatry frequently make. The first is to argue that advance directives are inimical to autonomy since they fail to put on a par a person's past and future selves. The second is to locate the significance of choice in the availability of options to choose from rather than the choice itself. Let us briefly consider each of them in turn.

The first move is key to critiques of pre-commitment, such as Jaworska's (see also Dresser 1995) which, by denying that temporal asymmetry has a crucial role to play in self-determination, are able to present pre-commitment as enforcing the will of a forceful 'then – self' upon a vulnerable 'now – self'. Following this line of thought, it seems plausible to consider pre-commitment as hostile to the autonomy of this 'now – self' which may have developed new interests and values, neither shared, nor anticipated by the 'then – self'. Yet, if applied to self-determination in general rather than specific psychiatric contexts, this thought becomes extremely implausible. For it would open any plan for the future to paternalistic interventions on grounds that one's future self may no longer stand for the same kind of thing as one's current self and the potentially different interests of the former have to be protected by society from the latter. In fact, a recent defence of hard paternalism draws on a similar line of reasoning in order to reach the conclusion that decisions of competent adults may be routinely ignored for their own good whenever the

---

wishes either to change in herself or to have done otherwise. The notion of externality introduced by Frankfurt (1999, Ch.5) provides a helpful contrast to trouble-free exercise of autonomy with respect to internal obstacles. See in particular the case of a person who struggles with unexpected jealousy at the news of a friend's success.

prospective intrusion will only violate their abstract right to self-determination but no specific and substantive rights, such as freedom of religion or association (Scoccia 2008, esp. pp. 371–379). This argument is instructive as it makes explicit the link between the claim that earlier ‘selves’ are not any weightier than later ‘selves’ (temporal neutrality) and the reduction of self-determination to a somewhat redundant confirmation of specific rights. As soon as we adopt the temporal neutrality thesis, self-determination loses both its object and purpose, for there is no longer a self to determine but only specific actions to undertake or abstain from. As a result, self-determination is reduced to mere freedom of action.

The second move is also related to the temporal neutrality thesis since its focus is the potentially irreversible character of pre-commitment in psychiatry. At the heart of this kind of critique lays the idea that, with respect to autonomy, pre-commitment is ambivalent an achievement to the extent that it forecloses options rather expands a person’s range of choices. This intuition underlies van Willigenburg and Delaere’s account of self-enforcement as incompatible with genuine autonomy. For this account implies that autonomy proper or self-legislation can only take place if there is no lapse of time to speak of between decision and action. More precisely, only if the decision to act is either made or confirmed at the time of action, can self-legislation be contrasted with self-enforcement as proposed by van Willigenburg and Delaere. This becomes apparent if we consider the demarcation point between the two: whether reasons for action are immediately present to the agent or in need of enforcement. By confining instances of self-legislation to contemporaneous actions, van Willigenburg and Delaere are able to interpret temporal asymmetry as a flaw of self-determination, typical of self-enforcement. Yet, the apparent contrast between the two boils down to focusing on one and the same phenomenon, choice, from different viewpoints in time. This is because irreversibility, for which self-enforcement is critiqued, is in fact a feature of successful agency over time. Implicitly, every plan that we bring to fruition in the present becomes a piece of self-enforcement in relation to future



actions. Moreover, the fact that options get foreclosed is the flipside of self-determination having taken place. To complain about pre-commitment for enabling courses of action that I may later regret is to fail to appreciate the significance of choice as key exercise of autonomy.

Aristotle's account of *proairesis* helps bring out this point.<sup>19</sup> Drawing on this account, we are able to observe that choice is an immediate cause of action which works by blocking alternatives. This is not to say that the chosen action gets carried out instantly, but that once an agent has made up her mind, there is no logical space for further deliberation either about other possible actions or the advantages of the action she already opted for.<sup>20</sup> So if a chosen course of action is not carried out, this can only mean that the agent has been prevented from following through by one kind of obstacle or another. Irreversibility and foreclosure of options turn out to be the signs of success at determining one's actions as opposed to having one's choices overturned. Moreover, the scenario, in which one's reasons for action are no longer apparent at the moment of action, looks like the standard case of acting for a reason rather than a paradoxical feature indicating a lesser kind of self-determination, as suggested by van Willigenburg and Delaere. Following this line of thought, self-legislation is at variance with pre-commitment only in so far it presents an aspiring kind of self-determination or self-determination which has not yet taken place. As soon as it achieves its objective, it becomes indistinguishable from self-enforcement.

Having dispelled the air of a paradox surrounding pre-commitment in psychiatry, we are able to draw two valuable lessons about autonomy and mental disorder. The first is as follows: not only are the three approaches to autonomy as set out at the start of the chapter conceptions of the same concept, there is an underlying hierarchy between them emerging from the discussion. More precisely, by considering key problems for critiques of pre-commitment in psychiatry, we are

---

<sup>19</sup> *Nicomachean Ethics* 3.1–5. See also Radoilska (2007, esp. pp. 211–232).

<sup>20</sup> If there is still room for deliberation, this only tells us that the agent has not yet made a choice. See Raz (1975) whose conception of exclusionary reasons expands on this line of reasoning.

able to establish that a value-neutral conception of autonomy in terms of authorship and ownership of one's motives and actions is central. In contrast, questions about reasons-responsiveness and social-relational status, which are defining respectively for a value-laden and a relational conception, become significant when either side of self-determination, authorship or ownership, no longer holds firm. For instance, by trying to elicit the reasons for a particular choice that looks unlikely to be autonomous, we get clearer about its authorship, that is, the extent to which it is up to a person rather than being the upshot of internal or internalised obstacles to her self-determination. To decide one way or the other, a spotlight has to be put on the putative reasonableness or choiceworthiness of the option taken. This is because, if this option can be asserted as something choiceworthy, over and above this person's professed preference for it, her authorship is clearly confirmed, despite the initial appearances.<sup>21</sup> Thus understood, choiceworthiness offers an additional guarantee that a putative choice is indeed a choice. This line of reasoning is well illustrated by the presumption of nonvoluntariness Joel Feinberg proposes to take as a starting point when considering cases of self-harm (1986, Ch. 20). The thought is not that self-harm can never be chosen voluntarily, but that there is need for further scrutiny. In a similar vein, we could say that responsiveness to reasons should come to the fore whenever the presence of internal obstacles to self-determination makes the assumption of non-autonomy more likely than not. The same argument applies, *mutatis mutandis*, to considerations of social-relational status in instances where the ownership of a choice requires closer examination. The idea is to ascertain whether the external obstacles present are so pervasive as to undermine active self-determination. In this respect, an assumption of non-autonomy could be thought of by analogy with the impact coercive offers may have on a recipient's circumstance of choice.

---

<sup>21</sup> See, however, Chapter 5 by Bortolotti et al. for an alternative account of the links between authorship and the ability to conceive one's choices as responsive to reasons.

The second lesson that we learn from looking into the paradox, to which pre-commitment in psychiatry seemed to give rise, is that there is a strong, though implicit assumption of non-autonomy that attaches to mental disorder. This would explain both the salience of value-laden and relational concerns in the critiques of advance directives and Ulysses arrangements I examined earlier and the intuitive appeal of these critiques. Moreover, since pre-commitment is a clear-cut, paradigm case of autonomy, the assumption at issue can plausibly target only the context of mental disorder in which pre-commitment takes place, not pre-commitment itself. How cogent is this assumption? To the extent that it is treated as defeasible, in continuity with the preceding reflection on cases, such as self-harm and coercive offers, the assumption of non-autonomy could be a good starting point in some instances of mental disorder, where outcomes of self-determination are not easily distinguished from obstacles to self-determination. However, drawing on the account of pre-commitment developed earlier, it is implausible to consider these instances as different in kind from standard cases of active self-determination. In this respect, we should be wary of a broad assumption of non-autonomy that attaches to mental disorder as such. More precisely, the worry is that by focusing on the gravity of obstacles to self-determination or the experiential unmanageability of mental disorder, we may be employing an unduly demanding but also misleading success criterion for autonomy, such as effortlessness.<sup>22</sup> As argued in this chapter, absence of obstacles is not a prerequisite for autonomy, nor is effortlessness an essential feature of its exercise. Failing to realise this, we shall run into a similar paradox to that faced by critics of pre-commitment in psychiatry who end up removing, instead of obstacles to self-determination linked to mental disorder, a vital means to get on top of them.

---

<sup>22</sup> On the notion of mental disorder as unmanageable distress, see Ch. 4 by Derek Bolton and Natalie Banner. See also Ch. 3 where K.W.M. Fulford and I reflect on the role of different success criteria in alternative accounts of mental disorder.

## **Summary and conclusions**

The structure of the concept of autonomy that emerged from the preceding discussion is as follows. Unlike the related freedom of action and intentional agency, autonomy is, firstly, incompatible with passive self-determination and, secondly, dependent upon a temporal asymmetry privileging prior over later commitments. More specifically, it takes the form of active self-determination with respect to one's actions, on the one hand, and, one's motives, on the other. There are two ways to exercise active self-determination: trouble-free autonomy and express pre-commitment. The effortlessness that distinguishes the former from the latter makes it difficult to perceive their shared form, which is pre-commitment. In contrast, this comes to light when active self-determination takes place against identifiable threats affecting either a person's authorship (internal obstacles) or ownership (external obstacles) over her actions and motives. The two paradigm kinds of express pre-commitment – trailblazing and character-building – articulate the underlying form, the first with respect to actions, the second with respect to motives.

In light of this analysis it is plausible to consider a value-neutral conception of autonomy as an independent source of normativity as more fundamental than value-laden and relational alternatives. This is because such a conception is focused on the complex relationship of authorship and ownership over one's actions and motives, which is at the heart of active self-determination. In contrast, considerations about responsiveness to reasons as opposed to mere incentives, which underpin value-laden conceptions, gain salience only when the presence of significant internal obstacles makes an assumption of non-autonomy plausible. Similarly, concerns about social and interpersonal standing central to relational conceptions, legitimately come to the fore when the external obstacles present are so overwhelming as to clearly back an assumption of non-autonomy. The latter two conceptions are therefore best understood as

offering each an additional test for distinguishing unobvious outcomes of active self-determination from apparent obstacles to it rather than rival theories of autonomy.

By making explicit the structure of the concept of autonomy, we are able to explain away the paradox to which express pre-commitment seems to give rise in the context of mental disorder. This paradox turns out to be the outcome of a flawed conceptualisation which takes effortlessness to be the form of autonomy, not active self-determination. As a result, express pre-commitment is downgraded to a secondary kind of autonomy. Moreover, irreversibility and the foreclosure of alternatives, key aspects of successful self-determination, are misconceived as undermining autonomy. By dispelling this misconception, it becomes clear that the obstacles to autonomy associated with mental disorder are not different in kind from the obstacles addressed by standard instances of pre-commitment. Consequently, there is no special case to be made for an assumption of non-autonomy attaching to mental disorder.

## References

- Aristotle. (1984). *Nicomachean Ethics*, trans. Ross, W.D. In Barnes, J. *The Complete Works of Aristotle*. Princeton, N.J.: Princeton University Press.
- Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. New York: Oxford University Press.
- Bolton, D. and Banner, N. (In press). Does mental disorder involve loss of personal autonomy? In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Bortolotti, L., Cox, R., Broome, M., and Mameli, M. (In press). Rationality and self-knowledge in delusion and confabulation: implications for autonomy as self-governance. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Bratman, M. (1999). *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

- Bratman, M. (2007). *Structures of Agency*. Oxford: Oxford University Press.
- Buss, S. (2008). Personal autonomy. In E. N. Zalta (ed.) *The Stanford Encyclopaedia of Philosophy* (Winter 2008 Edition), URL = <<http://plato.stanford.edu/entries/personal-autonomy/>>.
- Christman, J. (2003). Autonomy in Moral and Political Philosophy. In E. N. Zalta (ed.) *The Stanford Encyclopaedia of Philosophy* (Winter 2008 Edition), URL = <<http://plato.stanford.edu/entries/autonomy-moral/>>
- Christman, J. (2007). Autonomy, history, and the subject of justice. *Social Theory and Practice* 33(1): 1–26.
- Colburn, B. (2011). Autonomy and adaptive preferences. *Utilitas* 23(1): 52–71.
- Dresser, R. (1995). Dworkin on dementia: Elegant theory, questionable policy. *Hastings Center Report* 25(6): 32–38.
- Dworkin, R. (1993). *Life's Dominion: an Argument about Abortion and Euthanasia*. London: Harper Collins.
- Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press; Paris: Editions de la Maison des Sciences de l'Homme.
- Elster, J. (1984). *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.
- Elster, J. (2000). *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge: Cambridge University Press.
- Feinberg, J. (1980). *Rights, Justice and the Bounds of Liberty: Essays in Social Philosophy*. Princeton, Guildford: Princeton University Press.

- Feinberg, J. (1986). *The Moral Limits of the Criminal Law: Vol. 3 Harm to Self*. New York: Oxford University Press.
- Fistein, E. (In press). Conceptions of the good and the Mental Capacity Act. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68(1): 5–20.
- Frankfurt, H. (1998). *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.
- Frankfurt, H. (1999). *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.
- Frankfurt, H. (2002). Reply to Gary Watson. In Buss, S. and Overton, L. (eds.). *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge (Mass): MIT Press; 160–164.
- Fulford, K.W.M. and Radoilska, L. (In press). Three challenges from delusion for theories of autonomy. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Gallie, W.B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society* 56: 167–198.
- Gremmen, I. (2008). Ulysses arrangements in psychiatry: from normative ethics to empirical research and back. In Widdershoven, G., McMillan, J., Hope, T. and van der Scheer, L. (eds.). *Empirical Ethics in Psychiatry*; 171–185.
- Heal, J. (In press). Mental disorder and the value(s) of ‘autonomy’. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Hill, T. E. (1991). *Autonomy and Self-Respect*. Cambridge: Cambridge University Press.

- Holroyd, J. (In press). Clarifying capacity: value and reasons. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Jamison, K.J. (1995). *An Unquiet Mind: a Memoir of Moods and Madness*. London: Picador.
- Jaworska, A. (1999). Respecting the margins of agency: Alzheimer's patients and the capacity to value. *Philosophy and Public Affairs* 28(2): 105–138.
- Kant, I. (1785). *Groundwork of the Metaphysics of Morals*. In Kant, I. *Practical Philosophy*. (ed. and trans. Gregor, M.J.) (1996). Cambridge: Cambridge University Press.
- Korsgaard, C. M. (1996). *Sources of Normativity*. Cambridge: Cambridge University Press.
- Lillehammer, H. (In press). Autonomy, value and the first person. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Mill, J.S. (1859). *On Liberty*. Indianapolis (IN): Bobbs Merrill, 1959.
- Moran, R. (2001). *Authority and Estrangement: an Essay on Self-Knowledge*. Princeton (N.J.): Princeton University Press.
- Nordenfelt, L. (2007). *Rationality and Compulsion: Applying Action Theory to Psychiatry*. Oxford: Oxford University Press.
- Nozick, R. (1969). Coercion. In Morgenbesser, S., Suppes, P. and White, M. (eds.). *Philosophy, Science, and Method: Essays in Honour of Ernest Nagel*. New York: St. Martin's Press; 440–472.
- Oshana, M. (2006). *Personal Autonomy in Society*. Aldershot: Ashgate.
- Pettit, P. and Smith, M. (1996). Freedom in belief and desire. *Journal of Philosophy* 93(9): 429–449.
- Radden, J. (2005). Second thoughts: revoking decisions over one's future. *Philosophy and Phenomenological Research* 54(4): 787–801.
- Radoilska, L. (2007). *L'Actualité d'Aristote en morale*. Paris: Presses Universitaires de France.



- Radoilska, L. (In press). Introduction: personal autonomy, decisional capacity, and mental disorder. In Radoilska, L. (ed.) *Autonomy and Mental Disorder*. Oxford: Oxford University Press.
- Raz, J. (1975). Reasons for action, decisions and norms. *Mind* 84 (336): 481–499.
- Raz, J. (1987). Autonomy, toleration, and the harm principle. In Gavison, R. (ed.). *Issues in Contemporary Legal Philosophy: The Influence of H.L.A. Hart*. Oxford: Clarendon Press; 313–33.
- Scoccia, D. (2008). In defense of hard paternalism. *Law and Philosophy* 27: 351–381.
- Scanlon, T. (1996). The Difficulty of tolerance. In Heyd, D. (ed.), *Toleration: An Elusive Virtue*. Princeton: Princeton University press; 226–39.
- Stoljar, N. (2000). Autonomy and the feminist intuition. In Mackenzie, C. and Stoljar, N. (eds.), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*. New York: Oxford University Press; 94–111.
- Taylor, J.S. (2005). Introduction. In Taylor, J.S. (ed.). *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press; 1–29.
- Timmermann, J. (2009). Acting from duty: inclination, reason and moral worth. In Timmermann, J. (ed.), *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*. Cambridge: Cambridge University Press; 45–62.
- van Willigenburg, T. and Delaere, P.J.J. (2005). Protecting autonomy as authenticity using Ulysses contracts. *Journal of Medicine and Philosophy* 30(4): 395–409.
- Watson, G. (1975). Free agency. *Journal of Philosophy* 72(8): 205–220.
- Watson, G. (2004). *Agency and Answerability*. New York: Oxford University Press.

