

The Prudential Value of Autonomy

The principle of respect for autonomy is undeniably afforded particular salience in Western bioethics, and accounts of autonomy should aim to give an explanation as to why that is the case.¹ However, as well as seeking to give an account of the nature of autonomy's value, one might also question whether we ought to value autonomy so much, and how it should be weighed against other values. Our understanding of these issues will have significant implications for the many bioethical issues in which considerations of autonomy are invoked.

In this book, I have outlined a broadly Millian understanding of the nature of autonomy and its relationship to rationality. One might raise the concern that a Millian account is going to have trouble offering a satisfactory justification for a stringent requirement to respect autonomy that is consistent with Mill's broader utilitarian moral framework.² I shall comment on this particular interpretation of Mill below, but notwithstanding this issue, it is quite possible to claim that a Millian conception of autonomy can be adopted into moral frameworks that do not perfectly align with Mill's own. As such, my primary concern in this chapter is not how we can reconcile the value of autonomy within a broader consequentialist understanding of morality, but rather with how we should understand the value of autonomy itself. I shall argue that autonomy should be understood as not only instrumentally valuable, but also valuable for its own sake. The argument that I make for this claim has important implications not only for how we should weigh the value of autonomy against other values in bioethics, but also for how we should understand the nature of well-being.

At the outset, it is important to delimit the scope of my claims about autonomy's value in this chapter. It is sometimes claimed that autonomy has *moral* value, and that autonomy undergirds the moral value of personhood.³ On this approach, the principle of respect for autonomy can be understood as a particular instantiation of the more general moral principle of respect for persons. Modern statements of this view commonly find their source in Kant's moral philosophy, and his substantive account of autonomy.⁴ The moral respect due to a person on this approach reflects

¹ Walker, 'Medical Ethics Needs a New View of Autonomy', 595. For non-Western perspectives of autonomy's value, see Yang, 'Serve the People'; Kara, 'Applicability of the Principle of Respect for Autonomy'; Foster, *Choosing Life, Choosing Death*, 11.

² Walker, 'Medical Ethics Needs a New View of Autonomy', 603.

³ Kant, *Groundwork for the Metaphysics of Morals*, 4:435.

⁴ For some examples, see Velleman, 'A Right of Self-Termination?'; Darwall, 'The Value of Autonomy and Autonomy of the Will'; Habermas, *The Future of Human Nature*, particularly 37–44.

the high moral status that the person has as an autonomous agent, a being with intrinsic, non-exchangeable worth, or dignity that goes beyond mere price.⁵

As I pointed out in the introduction to this book, we may be sceptical about the extent to which this Kantian notion of autonomy is the sense that bioethicists typically intend to invoke in their discussions of autonomy. Whatever its merits, I shall not discuss it further here. As I have explained in previous chapters, my view of autonomy departs from Kant's substantive conception; accordingly, establishing that my procedural understanding of autonomy can provide a foundation for the moral value of personhood would require lengthy argument.⁶ Further whilst it is widely held that the value of autonomy has an important role to play in justifying the exercise of political power in liberal societies, I shall not be directly concerned with this question here.⁷ Instead, I shall focus my attention on whether autonomy bears prudential value; how, and to what extent does autonomy contribute to a person's well-being? I limit my discussion to this question in the hope that it has at least some bearing on other broader questions about the moral and political role of autonomy, on the assumption that the salient role of autonomy is at least partly attributable to its significant prudential value.

1. The Nature of Autonomy's Prudential Value

It is possible to distinguish two ways in which something can be prudentially valuable.⁸ Consider first, 'final value'. Something bears final value if it is valuable as an end, or for its own sake; for instance, knowledge, happiness, and virtue, *inter alia*, might plausibly be understood as bearing final value. We can contrast final value with 'instrumental value'; something has merely instrumental value if it is only valuable for the sake of something else.⁹ For instance, money has only instrumental value, in so far as it can be exchanged for other valuable goods.

Accordingly, if we are to claim that autonomy has instrumental value, we must also give an account of the valuable end to which autonomy serves as a means. *Prima facie*, one plausible candidate is well-being, broadly construed; a life lived

⁵ Kant, *Groundwork for the Metaphysics of Morals*, 4:435.

⁶ Jeff McMahan makes some remarks on this sort of project, and endorses the view that personal autonomy is a significant basis of the moral worth of persons. McMahan, *The Ethics of Killing*, 256–60. In a similar vein, Marilyn Friedman has argued that the first personal value of autonomy can provide reciprocity grounds for our moral obligations to others, and that personal autonomy is necessary for moral autonomy. Friedman, *Autonomy, Gender, Politics*, 60–7.

⁷ For a selection of relevant discussions of this topic, see Friedman, *Autonomy, Gender, Politics*, 75; Christman, *The Politics of Persons*; Raz, *The Morality of Freedom*; Mill, *On Liberty*; Spector, *Autonomy and Rights*.

⁸ For discussion of this distinction, see Korsgaard, 'Two Distinctions in Goodness'. Korsgaard's aim in this paper was to separate the distinction between final and instrumental value from the distinction between intrinsic and extrinsic value. The latter distinction pertains to whether or not something bears value in virtue of its intrinsic, non-relational properties, that is, 'in itself'. Although philosophers sometimes claim that autonomy has 'intrinsic' value, it seems that this is most naturally understood as the claim that autonomy has 'final' value. Whilst we may value autonomy as an end in itself, it is not clear that we value it by virtue of its non-relational properties. See Wall, *Liberalism, Perfectionism and Restraint*, 145 for discussion.

⁹ Korsgaard, 'Two Distinctions in Goodness', 170.

autonomously, it might be claimed, is more likely to lead to the attainment of the goods that make a person's life prudentially better. Following Parfit, theories of well-being are commonly classified into one of the following three types, as schematized below:

Hedonistic Theories—What would be best for someone is what would make their life happiest.

Desire-Fulfilment Theories—What would be best for someone is what, throughout their life, would best fulfil their desires.

Objective List Theories—Certain things are good or bad for us, whether or not we want to have the good things, or to avoid the bad things.¹⁰

We can also further distinguish *enumerative* theories of well-being from *explanatory* theories. The former sort of theory seeks to answer the question 'which things make someone's life go better for them?' In contrast *explanatory* theories of well-being seek to explain what it is about the things listed by enumerative theories of well-being that make them good for people.¹¹

The claim that autonomy is *only* instrumentally valuable is perhaps most congruent with explanatory hedonism; on such a theory, it might be claimed that autonomy makes a life go better just because autonomy is conducive to happiness (understood in terms of the experience of pleasurable mental states), which is the only thing that has final value on this view.¹² This understanding of the value of autonomy is commonly, although perhaps mistakenly, attributed to Mill.¹³ Such a reading of Mill might seem natural, given his insistence at the beginning of *On Liberty* that he regards utility as '... the ultimate appeal on all ethical questions' (a position that he defended in his *Utilitarianism*).¹⁴ Moreover, this understanding might seem plausible in view of the fact that individuals are in a privileged epistemic position with regards to the question of what will make them happy. As Mill puts the point:

With respect to his own feelings and circumstances, the most ordinary man/woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else.¹⁵

¹⁰ Parfit, *Reasons and Persons*, Appendix I. Although this tripartite classification is widely accepted, it has recently come under criticism, partly because it ignores Crisp's distinction between enumerative and explanatory theories. See Woodard, 'Classifying Theories of Welfare'. In the interests of clarity and space, I shall follow philosophical orthodoxy in discussing the tripartite classification, but I shall supplement this discussion with considerations pertaining to Crisp's distinction.

¹¹ Crisp, *Reasons and the Good*, 102–3.

¹² Happiness here is to be broadly understood in terms of the experience of pleasure (or desirable consciousness) and the absence of pain.

¹³ Robert Young also makes this observation in Young, 'The Value of Autonomy', 36. For examples of this interpretation of Mill, see Berlin, 'John Stuart Mill and the Ends of Life' and Ladenson, 'Mill's Conception of Individuality'. The problem with this interpretation is that it fails to acknowledge the way in which Mill departed from Bentham's monistic conception of utility. Mill's view actually seems to be that autonomy is incorporated into his understanding of utility.

¹⁴ Mill, *On Liberty*, 81; Mill, *Utilitarianism*.

¹⁵ Mill, *On Liberty*, 74. For a similar observation, see Feinberg, 'The Child's Right to an Open Future', 91. On the basis of these epistemic considerations, Dworkin refers to this view of the relationship between autonomy and well-being as the evidentiary view. Dworkin, 'Autonomy and the Demented Self', 7–8.

However, even if autonomy can be instrumentally valuable in this way, it is problematic to claim that it is valuable *only* in so far as it is a means to happiness. First, individuals will often be mistaken about what will make them happy; they may in fact achieve *less* happiness if they are left to their own autonomous devices than they would have done otherwise.¹⁶ To illustrate, we can imagine a young man who rejected his parent's advice and who autonomously decided that a career in finance would make him happy, but who comes to regret this decision in later life, when he realizes that he did not enjoy his career, and his choice meant forgoing a family life that he now believes his parents were right to suggest would have made him happy.

This observation alone is not an unimpeachable objection to the explanatory hedonist's claim that autonomy is only instrumentally valuable; perhaps most people *do* know what will make them happy, and counterexamples show only that there can be individuals whose autonomy lacks prudential value. In order to provide a stronger argument against the explanatory hedonist's claim, one would need to show that a life lived in the absence of autonomy could be *worse* than a life lived autonomously, even if the former life involved more happiness.

Consider an example in which this criterion is met. Would one believe that one's life would go better if one's affairs were to be determined by a wise and benevolent friend?¹⁷ Notably, this is something that Mill explicitly denies:

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, but because it is his own mode.¹⁸

More recently, James Griffin captures this Millian insight as follows:

... even if you convince me that, as my personal despot, you would produce more desirable consciousness for me than I do myself, I shall want to go on being my own master.¹⁹

Call the argument implicit in these observations the Personal Despot Argument (PDA). The thought underwriting the plausibility of the PDA is that autonomy has a special sort of value for us; there seems to be a value in *living a life of one's own* that is of central and fundamental importance to many of us.²⁰ Our rejection of even the wise and benevolent personal despot suggests that autonomy bears final value; we value autonomy for its own sake, and not just because we believe that being autonomous will lead to our attaining other prudentially valuable ends.²¹ On this

¹⁶ For a similar point, see Hart, *Law, Liberty and Morality*, 32; Dworkin, 'Autonomy and the Demented Self', 8; Hurka, 'Why Value Autonomy?', 364. Dworkin argues that Mill was also aware of this point. Dworkin, 'Paternalism', 73–4.

¹⁷ Wall, *Liberalism, Perfectionism and Restraint*, 146.

¹⁸ Mill, *On Liberty*, 131.

¹⁹ Griffin, *Well-Being*, 9.

²⁰ See also Wall, *Liberalism, Perfectionism and Restraint*, 129–30; Glover, *Causing Death and Saving Lives*, 96; Sher, *Beyond Neutrality*, 176; Kymlicka, *Liberalism, Community and Culture*, 12. The value that we tend to place on living a life of one's own offers a further clue as to why it seems problematic to claim that it is good, in a reason-implying sense, to live in accordance with an essentialist conception of an authentic self from which one feels alienated. The problem is that an agent who lives in accordance with such an alienated self does not seem to be engaged in a project of living a life of her own; rather, she is living a life of a self that she has *dis-owned*.

²¹ That the value here is *final* does not entail that it is not importantly related to other ends. For instance, Dworkin cashes out the value of autonomy by appealing to considerations of integrity. But the thought

approach, autonomy is at least partly *constitutive* of (rather than merely instrumental to) well-being.

I believe that the PDA captures an important truth, and I shall defend it in greater detail below. At this point though, we may observe that if the argument is indeed convincing, then it raises considerable problems for explanatory hedonism; however, both desire-fulfilment and objective list theories of well-being can accommodate our intuitive response to these examples, and allow for the view that autonomy has final value. Consider first desire-fulfilment theories; even if a personal despot could produce more happiness in your life, she would not be able to fulfil one's non-instrumental desire to live a life in which you make your own autonomous decisions. Alternatively, an objective list theory might simply claim that autonomy is an end that has final value. Indeed, many modern theorists have incorporated autonomy into their understanding of well-being in these ways.²² For instance, the desire for autonomy is a central desire in Griffin's own informed desire account,²³ and Sumner claims that well-being consists in '...authentic happiness, the happiness of an informed and autonomous subject'.²⁴ In a similar vein, John Finnis' description of the good of 'practical reasonableness' included in his objective list account seems to bear a close relation to autonomy as I have understood it in this book.²⁵

There are, however, important differences in how different theories of well-being account for the prudential value of autonomy. For instance, on enumerative *actual present* desire-fulfilment theories, autonomy is only incorporated into the good life for a particular person if they actually desire it. In contrast, on enumerative objective list theories that include autonomy, the final value of autonomy is not contingent upon the subject's desires in this way. I lack the space here to defend a full view of well-being. However, it should be acknowledged that the objectivist view of rational desires that I have defended in this book is based in part on a rejection of the view that our desires *simpliciter* can provide us with reasons. As such, the view of reasons, value, and autonomy that I have endorsed is incompatible with a purely desire-based *explanatory* account of well-being, since on such an account, the fact that something satisfies one of our desires makes that thing prudentially good for us; this sounds suspiciously like subjectivism about reasons of the sort that I rejected in Chapter 1. Accordingly, although the object-given view of reasons is compatible with a subjective desire-based account of well-being (as I discussed in Chapter 2), it is only so with respect to an *enumerative* desire-fulfilment account theory of well-being.

here is that living an autonomous life is constitutive of a living a life with integrity, rather than instrumental to it as a separate good.

²² Notice that a further benefit of incorporating autonomy into one's theory of well-being is that such theories are able to explain why the satisfaction of adaptive preferences may not enhance well-being. Sen raises this point in Sen, *Resources, Values and Development*, 304. See Sumner, *Welfare, Happiness, and Ethics*, 166 for discussion.

²³ Griffin, *Well-Being*, Part One, particularly 33–6.

²⁴ Sumner, *Welfare, Happiness, and Ethics*, 172.

²⁵ Finnis, *Natural Law and Natural Rights*, 88–90. Savulescu acknowledges the possibility of incorporating autonomy into an objective list account at Savulescu, 'Rational Desires and the Limitation of Life Sustaining Treatment', 213.

In view of the failure of explanatory hedonism to adequately capture the value of autonomy, and the incompatibility of purely desire-based explanatory accounts with objectivism about reasons, how should we understand the claim that autonomy has final value? The most plausible remaining strategy is to endorse an explanatory theory of well-being that appeals to objective values, and to claim that autonomy is one of the things that have such objective, final value.

However, autonomy should not be understood to be the *only* good on this sort of theory; one reason for this is that the realization of some values may require the absence of autonomy.²⁶ Furthermore, an adequate theory of well-being should allow for the possibility that pleasurable experiences can contribute in some way to well-being, even if one is not autonomous with respect to the choice to experience them. For instance, suppose that your affairs were determined by a benevolent personal despot, and they were incredibly successful in leading you to do things that led you to experience highly pleasurable mental states. We can still make sense of the claim that the pleasurable mental states you would experience would have *some* prudential value, even if the prudential value of the life would be severely impoverished by the absence of autonomy.

One might advance two related further claims here; first that the final value of autonomy is conditional on other components of the good life, and second, that autonomy may lack value or even be detrimental to well-being if it is put to bad uses.²⁷ There is a degree of truth in the first claim; on the theory of autonomy that I have developed here, there is an inextricable link between autonomy and the agent's values. Autonomy itself (and not its value per se) is conditional on the agent's beliefs about what constitutes the good life, since autonomous choices must be grounded in part by these beliefs.

In fact, it is also plausible to claim that something like the reverse relationship outlined in the first claim is true. Although certain goods (such as pleasure) are possible in the absence of autonomy, autonomy may plausibly be construed as a condition of other goods having a particular kind of value for the agent. It is through achieving the various objective values that partly constitute well-being through the *autonomous* pursuit of our own goals that we can understand ourselves as living a life that is *ours*. Whilst this need not be understood as either a necessary foundation of all other values, or even as something that is in fact universally valued, living a life that is one's own is prudentially valuable for its own sake. Only in a life in which the agent is autonomous with respect to the sustainment of the fundamental commitments that guide her conduct, and in her achievement of other objective goods, is it the agent herself who can be said to meaningfully *realize* the values instantiated in that life. It is this that is sorely absent in the life determined by a personal despot. Autonomy can thus be construed as being conditional to a particular kind of contribution that other goods make to well-being, one that serves to amplify the contribution a good makes in abstraction: that of contributing to a life that is meaningfully the agent's own.

²⁶ Berofsky, *Liberation from Self*, 248.

²⁷ Wall, *Liberalism, Perfectionism and Restraint*, 130; Varelius, 'The Value of Autonomy in Medical Ethics', 381.

The above reflections might be understood to be in tension with the second claim outlined above. Can autonomy have value (or have this amplifying effect on other values) even if it is put to immoral uses? I believe that the answer is 'yes'; to suppose otherwise is to confuse judgements about the all things considered goodness of states of affairs, with an assessment of what is good *for a person*.

To illustrate, suppose that Paul takes a great deal of pleasure from non-consensually harming people, and his violent actions help him to cultivate a self-narrative of himself as a dominant and powerful individual, a project that he takes to define his purpose in life. It is undoubtedly a terrible state of affairs that Paul performs these harmful actions. We may even plausibly say that his performing these actions autonomously *exacerbates* the badness of this state of affairs, and renders him more blameworthy for the harms caused. Finally, we may also justifiably restrain Paul from performing these actions by appealing to moral considerations that outweigh the value of his autonomy. However, I do not see a compelling reason to claim that it is worse *for* Paul that he performs these actions autonomously rather than non-autonomously. These are, recall, actions that he enjoys, and which he takes to be central to his character and his projects. If his autonomy in performing these actions makes Paul prudentially worse off, then it would have to be true that the performance of blameworthy immoral actions is detrimental to well-being, perhaps on the basis of the Aristotelian claim that moral virtue is a necessary constituent of well-being. However, this is a highly controversial claim that requires defence quite independently from considerations about the value of autonomy.²⁸

Another line of criticism to objective accounts of well-being that I have sketched here is that subjective attitudes seem to have an important influence on well-being, and it is not clear that objective accounts can accommodate this thought. All that matters for well-being on these theories is that objectively valuable things are incorporated into the agent's life; the agent's own subjective attitudes towards these goods are not important. Whilst there is considerable debate on the merits of this point, it suffices for my purposes here to say that if one finds this objection convincing, then it can be circumvented by adopting a hybrid account of well-being, according to which both the realization of objective values and one's holding subjective positive attitudes towards those objective values contribute to well-being. Such an account thus incorporates both objective and subjective elements.²⁹

The plausibility of such an account stems from the fact that although we may have reason to doubt that a theory of well-being that completely ignores individual preferences and attitudes is mistaken, it also seems plausible to claim that we can have self-interested reasons to want certain things, such as pleasurable experiences, loving relationships, and knowledge, even if we do not believe (perhaps incorrectly) that they will cause us happiness, or even if we do not desire them. Moreover, as

²⁸ For two arguments against this moralistic view of well-being, see Hurka, *Perfectionism*, 19–20 and Hooker's argument from sympathy in Hooker, 'The Elements of Well-Being', 25–7.

²⁹ Brad Hooker has recently defended a hybrid account that incorporates autonomy, and addresses theoretical questions about the limits of (and trade-offs between) subjective and objective elements in Hooker, 'The Elements of Well-Being'. See also Adams, *Finite and Infinite Goods*, 95–101; Feldman, *Pleasure and the Good Life*.

I explained in Chapter 2, the account of rationality that I have appealed to in this book allows for a degree of subjectivity, in so far as it is compatible with there being a plurality of goods, and with the possibility that rational agents can disagree about the weight that should be attributed to these different goods. Such disagreements do not arise simply because one party is wrong, and the other is right. Rather it is because of the imprecise nature of the truths governing many value comparisons.

2. Defending the Personal Despot Argument

As I discussed above, one of the main arguments in defence of the claim that autonomy bears final value is the PDA. However, this argument has been criticized on the grounds that it conflates the value that we attach to making decisions for ourselves with ‘... the value we attach to having our decisions reflect our deepest goals and values’.³⁰ To illustrate this, Mikhail Valdman suggests a thought experiment in which you have the opportunity to cede your final decision-making authority about how to act to a Personal Expert Committee (PC); this committee is better than you are at determining how to accomplish your goals and how to live according to your values.³¹ The PC, however crucially differs from the personal despot of the PDA. The PC does not determine your values; it only tells you how to live in accordance with them; the personal despot, on the other hand, might seek to increase your happiness by also harnessing control over your values.

Should we prefer the PC to self-government? Valdman suggests that we should, pointing out that we often cede decision-making authority in this way, such as in organizing our financial affairs.³² He also takes care to pre-emptively respond to a number of objections to his arguments.³³ Whilst I do not believe that all of these responses are satisfactory, I shall not pursue them here. Rather I shall raise two new objections to the PC argument. The first objection calls into question the scope of the PC example; the second objection suggests that, rather than showing that self-government has no intrinsic value, the PC example merely indicates that different elements of autonomy can have different value.

Let us consider the PC in a little more detail. Although the PC would intervene when it detected flawed practical reasoning, it would always use the agent’s own goals and values as the basis for its decisions. To illustrate, suppose that David has some prudential goal *X*, and has to choose between two possible acts *A* and *B*. Suppose that out of these two acts, only *A* would serve as a reliable means to David’s achieving *X*; in this case, the PC would only intervene if David believed that he prudentially ought to *B*.

Valdman’s model is not problematic when the value of the goal (*X*) is distinguishable from the acts that one must perform as a means to achieving that goal. However, it is problematic when this is not the case. Yet, the value of a number of goals is inextricably related to the way in which we achieve that goal. To illustrate, suppose that one valued being able to play a complex piece of music on the piano, say

³⁰ Valdman, ‘Outsourcing Self-Government’, 764.

³¹ *Ibid.*, 770.

³² *Ibid.*, 772.

³³ *Ibid.*, 780–9.

Rachmaninov's second concerto. In a crude sense, in order to play this piece, one would simply need to hit certain combinations of keys, in a certain order, for a certain time. In order to play the piece in this crude sense, one would need to develop excellent motor skills and technique, normally through devoting hours to practising the requisite movements, and to learning the structure of the piece. Whilst it might be claimed that there is some value in the discipline and effort that this practice requires, the goal of being able to play Rachmaninov's concerto in the crude sense under consideration could retain its value for an agent, even if they achieved it via a more efficient means that did not involve effort or discipline; for example, instead of sitting through hours of lessons and practice sessions, suppose (somewhat fantastically) that one could simply 'download' the ability to play the right notes in the right order for the right amount of time.

In this crude sense of being able to play the piece, the value of the goal is distinguishable from the means that one takes to achieve it. However, consider now someone who has a more refined desire to be able to play the Rachmaninov piece; rather than valuing being able to simply 'play the right notes', this person values being able to play the piece according to their own interpretation of the music. This might involve, *inter alia*, their deciding which phrases of the piece need particular emphasis, and the strength they should exert in pressing the keys at particular points. Whilst the achievement of this goal requires the same abilities as the goal of playing the piece in the crude sense, it also requires something more, something like *creativity*; and because of this, it seems that the value of the goal is inextricably linked to the fact that the agent herself exercises *her own* creativity in its pursuit.

This is important, since on this understanding of the value of the goal, it does not make sense to say that one might be able to better achieve the goal by outsourcing to something like a PC. A PC, an expert tutor, or a futuristic downloadable music program could make you a better technical piano player; and this technical ability might be prerequisite for going on to exercise one's creativity in playing. However, completely relying on a PC to realize the goal of playing Rachmaninov's second concerto in a sophisticated sense would defeat the value of the goal itself.

The point that this example raises is that the relationship between the value of our goals and the means that we take to achieve them is not always as simple as the PC argument implies. Whilst Valdman is correct to point out that we often outsource decision-making authority, the examples he highlights are cases in which the value of the goal is clearly distinguishable from the manner in which the goal is achieved; for example, the value we attribute to achieving financial security is rarely taken away if we attain it by allowing a financial adviser to make our financial decisions for us.³⁴ However, in more complex cases, the value of some goals seems to be at least partly dependent on the fact that in achieving the goal, the agent herself makes her own mark in doing so. Rob Goodman captures a similar thought in his distinction between 'process goods' that pertain to excellence in the performance of an activity,

³⁴ *Ibid.*, 772.

and ‘outcome goods’, that pertain to the benefits that an activity creates.³⁵ Playing Rachmaninov in the crude sense would qualify as an outcome good in my example, whilst playing the piece in the sophisticated sense would involve process goods.

The above considerations lend support to the claim that simply ensuring that an agent’s goals are achieved may not facilitate their autonomy. The fact that some of our goals are process goods lends support to the claim that, at least in some cases, a ‘... crucial part of the notion of “self rule” is that it is *me* that achieves my goals’.³⁶ Although it seems that many of the goals that agents tend to have involve process goods, let us suppose that the PC would not intervene to ensure the achievement of process goods, and that the objection still stands in relation to a number of other goals that people tend to have. Even if we concede this point, the objection only shows that the value of different sorts of autonomy can come into conflict, not that autonomy lacks final value.

There are two central points undergirding this line of response. The first is that according to the PC argument, one may fail to be self-governing even when one is living in accordance with one’s own goals and values. According to the terms of the argument, one will fail to be self-governing if it is the PC rather than the agent herself who ensures that they live in accordance with their goals and values. The second point concerns the distinction that I have drawn upon in this book between global and local autonomy. As I explained in the introduction, we can understand autonomy to be a property of agents in a particular time-slice, with respect to a particular decision. When we conceive of autonomy in this way, we are considering *local* autonomy. In contrast, we can also understand autonomy as a *global* property that agents can instantiate diachronically.

Notice that when the PC argument stipulates that one may fail to be self-governing even if one is living in accordance with one’s own goals and values, the failure here is a failure of local, rather than global autonomy. After all, *ex hypothesi*, the PC would only govern you in accordance with your own deeply held commitments and values. As such, the PC will only intervene when one’s own *local* decision-making is likely to prove counter-productive to one’s pursuit of the long-term goals that may be understood to undergird one’s *global* autonomy.

The reason that the PC argument may appear to be convincing is that it fails to adequately distinguish local and global autonomy. Although it might be true that there are cases in which we could have good reason to outsource our decision-making to experts, the strength of this reason is *itself* rooted in the value of being able to live what Valdman calls an ‘acceptable’ life, in accordance with one’s own freely chosen goals and values;³⁷ however, this is simply what it is to be *globally* autonomous. Accordingly, the PC argument is only sufficient for proving that the value of local and global autonomy may sometimes be in conflict, and that we would

³⁵ Goodman, ‘Cognitive Enhancement, Cheating, and Accomplishment’, 146 and 152–4.

³⁶ Sandman and Munthe, ‘Shared Decision Making, Paternalism and Patient Choice’, 66. These authors make the strong claim that this is always true of respect for autonomy. However, I limit my endorsement of this claim, as my discussion below shall clarify.

³⁷ Valdman, ‘Outsourcing Self-Government’, 769.

often prioritize our global autonomy over our local autonomy. Yet, this is not a problematic conclusion for those who claim that autonomy bears final value.

3. The Value of Different Elements of Autonomy

The second response to the PC objection turns on the claim that it is possible for local autonomy to come into conflict with global autonomy. On some views of the relationship between local and global autonomy, this claim would be implausible. For instance, it would be implausible if one held the view that global autonomy is simply an aggregate of the instances of local autonomy over time.³⁸ However, in the introduction, I suggested an alternative understanding of the relationship, according to which an agent's global autonomy depends on the extent to which she lives in accordance with her own diachronic plans and commitments. Ronald Dworkin also implicitly acknowledges that the value of global autonomy is distinct from the value of individual decisions in outlining his integrity view of the value of autonomy as follows:

[A]utonomy makes each of us responsible for shaping his own life according to some coherent and distinctive sense of character . . . This view of autonomy focuses not on individual decisions one by one, but the place of each decision in a more general program or picture of life the agent is creating . . .³⁹

Despite this, in some cases, being able to make our own local choices *is* essential to the facilitation of our global autonomy. This is not simply because of our privileged epistemic access to knowing which goals we value; rather it is because the goals we aim to pursue in some cases are process, rather than outcome goods. Making local decisions about how one pursues such goals is inextricably linked to one's evaluation of the achievement of the goal itself.

However, when we consider the pursuit of outcome goods, conflicts between local and global autonomy become far more acute. In such cases, the value of one's goal can be distinguished from the value of making locally autonomous decisions about how to pursue it. Indeed, it may even be the case that we could better facilitate an agent's pursuit of the goal that undergirds their global autonomy by restricting their local autonomy. Which of these elements of autonomous agency should we prioritize?

In advocating the PC objection, Valdman himself implicitly highlights one possible explanation of why global autonomy might have precedence over local autonomy. As Valdman suggests, the deep commitments that one must live in accordance with to live an 'acceptable' life are central to our identity, on psychological understandings of that concept. This is not true of many of the short-term goals that our local decision-making concerns. These may sometimes be trivial, and in no way connected to any of our deep global commitments; I can, for instance, be locally

³⁸ Christman, 'Autonomy and Personal History', 18–19.

³⁹ Dworkin, 'Autonomy and the Demented Self', 8; see also Dworkin, *Life's Dominion*, 224. Interestingly, as Foster notes, judges invoked Dworkin's understanding of the value of autonomy in their judgement on *Chester v. Afshar*. See Foster, *Choosing Life, Choosing Death*, 84.

autonomous with respect to my decision about what to have for lunch. In contrast, the adoption of a long-term goal requires a far more significant kind of commitment; in making decisions about such commitments, we can clarify and shape the nexus of our judgements about what is valuable. As I discussed in Chapter 2, such judgements play a highly significant role in our character systems. As such, when the two sorts of autonomy cannot both be realized, concerns pertaining to our sense of identity, of defining who we are, may give us reason to prioritize our global autonomy over our local autonomy.

Accordingly, it should not be surprising that patients often believe that the best way to achieve their global commitments in a medical context is to sacrifice their local autonomy with respect to their treatment decisions by telling their doctor to 'do what you think would be best'. This should be viewed as an expression, rather than an abdication, of autonomy.⁴⁰ For instance, the patient might not trust herself to make a difficult local decision that is in harmony with her evaluative judgements, or she may not feel able to weigh the complex information involved in such a decision appropriately. Crucially, in light of my response to the PC objection, the amelioration of a patient's condition is typically most naturally understood as an outcome good, rather than a process good; as such, the patient may outsource her decision-making here without this undermining the value of her diachronic goal. Accordingly, an agent may retain her global autonomy in making this request if the patient believes that the doctor is more likely than she is to make a treatment decision that would best reflect her own evaluative judgement about what would be good for her in a reason-implying sense.

In light of this discussion of the significance of global autonomy, one might be tempted to ask why we should worry about locally autonomous decisions at all, rather than simply focusing on global autonomy alone. There are two reasons for why we should resist this temptation. First, as I discussed above, many of the goals that undergird our global autonomy are process goods. Second, as I explored in Chapter 6, an individual's local autonomous decisions can have considerable moral significance when they concern whether the agent wishes to exercise their power to waive a claim right. Although I argued that local autonomy has some role to play in our understanding and justification of rights such as these, the interest they serve to protect may crucially not be best understood in terms of their contribution to an individual's global autonomy. To take Archard's example again, if one violates another's right to bodily integrity by non-consensually inserting a painless mouth-swab, the wrong done here is not plausibly construed as one of subverting the victim's ability to lead her life as she chooses.⁴¹

It is far from clear that considerations of the agent's own global autonomy are sufficient to justify the infringement in this case.⁴² More broadly, an autonomous

⁴⁰ In view of my discussion here, we should not view the empirical data concerning patient behaviour that Foster presents on this point as undermining the significance of autonomy in the manner that he intimates. See Foster, *Choosing Life, Choosing Death*, 97.

⁴¹ Archard, 'Informed Consent', 22.

⁴² For more on the distinction between right violation and infringement, see Thomson, *The Realm of Rights*, 82–104.

agent can plausibly wish to exercise her autonomy by refusing to waive a right against some interference, even when that interference would facilitate the pursuit of goals that undergird her global autonomy. The interest it protects may, in some cases, have greater reason-giving strength for an individual.

Despite these caveats, we should not be blind to the cases in which there can be conflicts between global and local autonomy in which considerations of global autonomy might plausibly win out. When the goals in question concern outcome goods, and facilitating the agent's pursuit of those goals would not require interference of the sort that would violate a powerful claim right, we might plausibly be justified in prioritizing the agent's global autonomy over her local autonomy. To unreflectively assume that appropriate respect for autonomy *always* demands that we should respect locally autonomous decisions that will certainly undermine the agent's pursuit of what we know to be the goals that undergird her global autonomy, is to fetishize one kind of autonomy over another, in a manner that does not reflect the prudential value of these different kinds of autonomy. Respect for autonomy should thus consider both global and local understandings of that which is being respected.

The above considerations lend an autonomy-based justification of weak paternalism. According to weak paternalism it is legitimate to interfere with the *means* that agents choose to achieve their ends, if those means are likely to defeat their own ends. For example, suppose Fred is overweight and autonomously wants to lose weight; however, he chooses means to this end that will not be effective, buying unproven weight-loss products he finds on the internet, continuing to eat unhealthy food that hinders his ability to lose weight, and refusing to exercise. The weak paternalist would claim that we could intervene in order to ensure that David will choose a more effective means of reaching his goal, perhaps by restricting the availability of unhealthy foods. On the autonomy-based approach I am outlining, weak paternalist measures could be justifiable in such cases if the goal in question is an outcome good, and the measures would not infringe upon a claim right that David has not waived. Similarly, the above considerations offer the most plausible prospect of a broadly autonomy-based justification of nudges that undermine local autonomy. Such a justification would qualify as a weak form of paternalism outlined above.⁴³

Weak paternalism can be contrasted with strong paternalism, which states that it is legitimate to interfere to prevent people from achieving those ends that they are mistaken in believing to be good for them.⁴⁴ For instance, suppose that Grant is overweight but believes that this is not something he ought to worry about; in fact, Grant values the experience of gastronomic pleasures over his health, and accepts the health risks that his lifestyle involves. A strong paternalist might potentially claim that Grant is weighing his values incorrectly here, and that it may be permissible to somehow restrict Grant's intake of unhealthy foods. Whilst strong paternalism requires a particular kind of beneficence-based justification (as I shall explore below), weak paternalism may be justified by an appeal to the precedence of global

⁴³ For a similar strategy in favour of limiting informed consent procedures, see Levy, 'Forced to Be Free?'

⁴⁴ Dworkin, 'Paternalism'.

over local autonomy, in so far as it calls for the safeguarding of the agent's ability to effectively pursue their own ends, over their freedom to make locally autonomous decisions about the means to take to their end.

It might be claimed that in advocating weak paternalism, I am betraying my above comments about the prudential value of autonomy; indeed, Sarah Conly's recent book defending a broadly similar strategy was titled *Against Autonomy*.⁴⁵ I cannot deny that there is a sense in which the approach that I am outlining here is 'against autonomy'. However, the point of this discussion has been that because local and global autonomy can come into conflict, we *have to* be against autonomy in one of these senses if we are to reconcile these conflicts. My argument has been that we should prioritize the kind of autonomy that often plausibly holds more significant prudential value, and it is a mistake to believe that this will always be local autonomy.

However, a key point underlying this justification is that the use of weak paternalism must be limited to cases in which it will promote what we know to be the values and goals undergirding the target's global autonomy. Given epistemological barriers to knowing that this will be the case,⁴⁶ the blunt nature of most proposed paternalist interventions, and the fact that reasonable agents can disagree about the weight they attribute to different goods, the scope of justifiable paternalist strategies on this approach will be extremely limited. Indeed, as I have argued elsewhere, we should be concerned about the possibility that these features of the justification might be overlooked, and the back-door perfectionism that could be ushered in under the guise of such weak paternalist justifications of manipulative interferences.⁴⁷ This would amount to considerations of beneficence (narrowly conceived in the sense that I discuss below) being dressed up in the language of (global) autonomy.

To conclude the discussion of the value of different elements of autonomy, could the decisional and practical dimensions of autonomy be valuable to different degrees? I have argued that the decisional element of autonomy is theoretically prior to the practical dimension. This analysis might tempt one to claim that whilst the decisional element of autonomy might bear final value, the practical dimension of autonomy might bear only instrumental value.⁴⁸ However, this thought should be resisted, since it fails to adequately capture the point that the way in which we value autonomy for its own sake is, as I suggested above, inextricably related to our fundamental interest in *living* a life that is our own, in acting on the basis of our autonomous decisions. Even if it were possible to separate the two dimensions of autonomy into discrete categories (which, I argued in Chapter 5, is doubtful) neither dimension alone seems sufficient for the project of living a life of one's own. This point is perhaps clearest with respect to practical autonomy; the fact that an agent is able to act effectively in pursuit of an end that she does *not* autonomously desire does not seem to be valuable

⁴⁵ Conly, *Against Autonomy*.

⁴⁶ For this reason, the relationship that the individual performing a weakly paternalistic intervention bears to its recipient can have considerable implications for its permissibility. For the significance of relationships to understanding permissible manipulative interference, see Blumenthal-Barby, 'A Framework for Assessing the Moral Status of "Manipulation"'.
⁴⁷ Pugh, 'Coercive Paternalism and Back-Door Perfectionism'.

⁴⁸ In a similar vein, Taylor claims that increasing an agent's freedoms does not increase her autonomy, but rather increases the value of her autonomy. Taylor, *Practical Autonomy and Bioethics*, 6.

in the same way as the ability to act in pursuit of an end that one autonomously desires;⁴⁹ only then can this freedom be said to be integral to the agent's ability to lay out her own mode of existence.

The value of decisional autonomy is also to a considerable extent conditional on the presence of practical autonomy, although this point is perhaps less immediately obvious. The reason for this is that it is difficult to imagine cases in which an agent lacks *all* freedoms that are relevant to their practical autonomy. To illustrate, reconsider the case of the slave philosopher Epictetus. Despite his being enslaved, and thus seemingly lacking *any* freedom, one might claim that Epictetus nonetheless represents the epitome of autonomy, in so far as he defied his lack of freedom by spending his life in the pursuit of a self-determined goal; namely the pursuit of philosophical truth. However, this example does not demonstrate that decisional autonomy *alone* is valuable for its own sake. Although Epictetus lacked many practical freedoms, he crucially retained the freedom to act effectively in pursuit of his goal of philosophizing; to this extent, he had both decisional and practical autonomy.

Therefore we should reject the claim that one dimension of autonomy is more prudentially valuable than the other. Neither dimension of autonomy in abstraction from the other is prudentially valuable for its own sake. Rather, we should understand the conjunction of the two dimensions of autonomy to form an organic whole which is prudentially valuable for its own sake, and whose value is derived from fundamental value in the exercise of laying out our own mode of existence, of living a life that is our own. It is not enough to be autonomous in our practical deliberations; we must also be able to act on the basis of those deliberations. This, I take it, is part of what John Harris means to capture in his bold claim (quoted in Chapter 5), that 'without agency, decision-making is . . . both morally and practically barren'.⁵⁰

4. Autonomy and Conflicting Values in Bioethics

Prior to outlining the sorts of values with which autonomy might come into conflict, it is crucial to first establish that autonomy can conceptually come into conflict with other values at all. On some views, autonomy cannot come into conflict with other values, because autonomy itself is understood to be the source of all other values.⁵¹

I mention this so-called 'autonomism' view only to reject it. I do so for the following reasons: First, there are some values to which autonomy cannot plausibly serve as a source, because their very possibility presupposes the absence of autonomy. For instance, certain values such as familial unity and dignity can be found in

⁴⁹ For this reason, I am less concerned than Berlin about unavoidable conflicts between his conceptions of positive and negative liberty, although I similarly acknowledge the potential for abuse of the prospect of constraining negative liberty in the name of positive liberty. See Berlin, 'Two Concepts of Liberty', 44–5. Further, I am not denying that practical freedoms can have other forms of instrumental value; the effective pursuit of non-autonomously endorsed goals could lead to other goods such as pleasure. Contrary to what I have claimed here, Feinberg has argued that freedom has intrinsic value. See Feinberg, *The Moral Limits of the Criminal Law*, 211–12. For a comprehensive rebuttal of these arguments see Haworth, *Autonomy*, 139–47. See also Griffin, *Well-Being*, 237.

⁵⁰ Harris, '“ . . . How Narrow the Strait!”', 249.

⁵¹ Haworth, *Autonomy*, 7 and 184.

communities that eschew autonomy, and the presence of these values in such communities is ‘...in part a function of the very absence of individual autonomy’.⁵² Second, it seems that we can make sense of a life incorporating prudential value even if it lacks autonomy. As I argued above, whilst we may disagree with the hedonist’s claim that we should hand over control of our lives to a benevolent personal despot if we had the chance, this does not entail that such a life would lack *any* prudential value; *ex hypothesi*, it would contain a great deal of pleasure. It seems implausible to claim that this pleasure would count for *nothing* simply because the agent in question lacked autonomy. Finally, we will clearly need further moral principles to guide us in cases of competing autonomy claims, when satisfying the autonomous preference of an individual requires frustrating the autonomous preferences of another.

It would thus be ‘absurdly simplistic’⁵³ to understand autonomy to be the sole value governing medical ethics. However, the truth in autonomism is that autonomy seems to be related to a particular sort of value that we understand to be salient; in section 1, I described this as the value of ‘living a life of one’s own’. For this reason, although autonomy can conceptually come into conflict with other values in bioethics, one might maintain that autonomy is likely to win out in such conflicts. Indeed, one might worry that this understanding of the value of autonomy lends support to what Onora O’Neill disparagingly calls the ‘consumerist view’ of autonomy, according to which considerations of respect for autonomy serve as both necessary and sufficient conditions for the moral justification of some course of action.⁵⁴ In the following discussion, I shall explain that this view is neither reflected in medical law, nor a corollary of the understanding of the value autonomy I have outlined here.

According to the widely invoked four principles approach to biomedical ethics, our ethical decision-making should be governed by four ethical principles; namely, the principle of beneficence, the principle of non-maleficence, the principle of autonomy, and the principle of justice.⁵⁵ In outlining these principles, Beauchamp and Childress explicitly claim that none of these principles takes priority over any of the others; as such, it would be a mistake to assume that autonomy should trump these other values.⁵⁶

One of the clearest examples of where the consumerist understanding of autonomy fails is in the context of health resource allocation, where considerations of autonomy and justice will often conflict. Since the demand for many health resources (such as organs for transplantation and hospital ward space) far outstrips supply, it is not the case that the autonomous wishes of all the patients who wish to use these

⁵² See Berofsky, *Liberation from Self*, 248. Oshana also posits ‘security’ as a value that can conflict with autonomy. See Oshana, ‘How Much Should We Value Autonomy?’, 113–14.

⁵³ Foster, *Choosing Life, Choosing Death*, 9.

⁵⁴ O’Neill, *Autonomy and Trust in Bioethics*, 2.7 particularly p. 47.

⁵⁵ Beauchamp and Childress, *Principles of Biomedical Ethics*. In his discussion, Foster suggests that medical law and ethics should also consider principles relating to professional integrity and rights and duties of doctors and patients. See Foster, *Choosing Life, Choosing Death*, ch. 2.

⁵⁶ See Beauchamp and Childress, *Principles of Biomedical Ethics*, especially 57 and 177. However, it is questionable whether all those who invoke the four principles approach abide by this dictum; see Gillon, ‘Ethics Needs Principles—Four Can Encompass the Rest—and Respect for Autonomy Should Be “First among Equals”’.

resources can be respected. Indeed, when societies have to make decisions about health resource allocation, considerations of distributive justice will unavoidably take precedence over considerations of individual autonomy, since it is simply not possible to satisfy the autonomous preferences of each person to have access to the scarce resource.⁵⁷ Since individual patients cannot generally be said to have a positive claim right to a scarce medical resource, an autonomous request for treatment generates a much weaker moral reason than a patient's autonomous decision not to waive negative rights that she does enjoy, when she refuses medical treatment.⁵⁸ In any case, contrary to what the consumerist view implies, reasons of autonomy are often not sufficient to justify actions, including the allocation of a scarce resource to an individual, given the implications that such actions can have for others.

This is quite compatible with the claims I have advanced in this chapter. Even if we accept the view that autonomy is fundamental to an individual's well-being, this is compatible with the claim that moral reasons generated by the well-being (and indeed the autonomy) of *others* can be sufficient to outweigh our reasons to respect the autonomy of the individual herself. Despite his staunch defence of liberty and individuality, even Mill claimed that considerations of justice can override the individual's right to liberty in this way. This thought is apparent in his 'Harm Principle', according to which:

The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.⁵⁹

Strikingly, the Harm Principle allows for the possibility that an individual's negative rights can, in some cases, be permissibly infringed by Mill's own lights; respect for autonomy then is not necessary for the moral justification of some actions, contrary to the consumerist view. Of course, one broadly autonomy-based reason that might justify a restriction of liberty against an individual's (occurrent) will is if the individual can be understood to have implicitly consented to a particular infringement, on the basis that doing so is a condition of the social contract that affords them a number of other strong and important protections. For instance, one might plausibly explain why one may justifiably prevent a would-be thief from robbing a bank in this way, even if one believes that the thief's autonomy would be best served by both affording him these protections *and* the freedom to rob the bank.

However, other restrictions of liberty might be justified by the need to safeguard the interests of others. Infectious disease control is an example where one might plausibly exercise power over another person against their will in a manner that is

⁵⁷ Pugh, 'Navigating Individual and Collective Interests in Medical Ethics'.

⁵⁸ If individuals have a positive claim right to a particular medical treatment, this would entail that physicians have a duty to provide it. We may also note that this might generate reasons that speak against an autonomy-based positive right to treatment. For instance, one might contend that such a right would compete with doctors' right to conscientious objection. Foster, *Choosing Life, Choosing Death*, ch. 8; Schuklenk and Smalling, 'Why Medical Professionals Have No Moral Claim to Conscientious Objection Accommodation in Liberal Democracies'; Cowley, 'Conscientious Objection in Healthcare and the Duty to Refer'; Wicclair, 'Justifying Conscience Clauses'. Alternatively, one might hold that doctors do not have a duty to perform treatments that violate unwaivable claim rights.

⁵⁹ Mill, *On Liberty*, 80.

morally permissible by the lights of the Harm Principle. Suppose a person infected with the Ebola virus refused to enter isolation voluntarily, and thus risked spreading the virus to other members of his community; it seems plausible to claim that public health authorities would be ethically justified in enforcing compulsory isolation on such an individual. In England and Wales, the right of the state to impose compulsory isolation in this sort of situation is legally enshrined in the Health and Social Care Act 2008.⁶⁰

Contrary to the consumerist view, respecting autonomy thus does not serve as a necessary condition of all kinds of permissible medical intervention. However, simply invoking the language of autonomy alone to understand the forms of protection that are owed to individuals in public health contexts is perhaps too blunt an instrument; depending on one's overall approach to ethics, we may also need to take into account the degree of harm that an intervention will cause (on broadly consequentialist approaches) or the kinds of rights it would violate (on broadly deontological approaches). To illustrate the importance of this, suppose that instead of refusing isolation, an individual refused to undergo an invasive medical injection that was necessary to prevent him from spreading a deadly infectious disease to others. It is less straightforwardly clear that it would be permissible to impose this treatment; indeed, in England and Wales, the law allowing the imposition of quarantine explicitly rules out the imposition of non-consensual treatment, including vaccinations.⁶¹ However, in both cases, we may presume that the moral reasons generated by considerations of public safety are held constant.

In order to make sense of the intuition that the individual has a stronger claim against the imposition of a medical treatment than he does against the imposition of quarantine, it seems that one would have to supplement considerations of the individual autonomy of the quarantined individual and harm to others with considerations of the quarantined individual's rights, or the harms that non-consensual quarantine will do to them. For instance, on a rights-based approach, it might be claimed that whilst quarantine violates the individual's right to freedom of movement and association, non-consensual medical treatment violates the recipient's right to bodily integrity. It might then be argued that the latter is a more robust right.⁶² Alternatively, it might be claimed that bodily invasive interventions cause greater harms than placing restrictions on an individual's freedom of movement and association.

As I suggested in previous chapters, it is a mistake to think that the justification of claim rights that incorporate a power to waive the claim can be entirely divorced from considerations of autonomy. In a similar vein, we may note that the harmfulness of an intervention may plausibly turn to some degree on the strength of the individual's autonomous preference to avoid that interference. So these considerations are importantly related to autonomy. However, in order to adequately capture

⁶⁰ Health and Social Care Act 2008, Part 2A, 45.

⁶¹ Ibid. Although see Herring, *Medical Law and Ethics*, 169 for discussion of courts' reluctance to preclude the permissibility of non-consensual treatment of competent individuals to save others.

⁶² The UK Health and Social Care Act implicitly seems to endorse this view, in so far as it permits non-consensual quarantine, but prohibits non-consensual medical treatment.

the nuances of whether public safety should outweigh considerations of autonomy, we should also take into account these further considerations. The issue cannot be settled just by establishing whether the agent has or has not autonomously consented to the intervention itself.

Furthermore, the nature and degree of the harms that we can expect to *prevent* by violating the individual's autonomy should also factor into our overall weighting of these competing values. In the case of preventing the spread of a deadly pandemic, we might claim that we may be permitted to impose a very invasive non-consensual intervention that would prevent the spread of disease, given the number of lives at stake; in contrast, many find abhorrent the possibility that it might be permissible to carry out non-consensual interventions in order to save very few lives. Consider for instance the permissibility of carrying out a Caesarean section that is necessary to save an unborn child's life on a non-consenting competent woman in labour.⁶³

The Harm Principle claims that it can be permissible to intervene on an individual's liberty to prevent the non-consensual imposition of harms to third parties that might otherwise occur. However, it also states that this is the *only* justification for such an intrusion on the individual; indeed, the main thrust of the Harm Principle is that the principle of autonomy trumps considerations of individual beneficence. This asymmetry between the moral weight of harm to others and harm to self in conflicts with autonomy may be understood to be a reflection of the fact that the individual herself may be understood to tacitly consent to harm when they autonomously choose to engage in self-harming behaviour, whilst third parties do not similarly consent to being harmed.⁶⁴

Mill himself limited the application of the principle to those 'human beings who are in the maturity of their faculties'.⁶⁵ So, the fact that it can be lawful to perform unwanted beneficial medical procedures on patients who lack decision-making capacity is quite compatible with the principle. However, even observing this feature of Mill's thought, the salience attributed to autonomy on the Millian approach is only *partially* reflected in medical law. Contrary to those who perceive a consumerist view of autonomy at work in modern bioethics, elements of medical law are quite difficult to reconcile with the Millian understanding of the competing values of autonomy and well-being. As I explored in Chapter 6, consent cannot serve as a sufficient defence for certain kinds of intervention; accordingly, it seems that individuals may plausibly be said to enjoy certain negative claim rights that are *not* attended by the second-order power to waive those claims. Since such powers significantly demarcate the individual's sphere of autonomy in the law, it seems that the absence

⁶³ For discussion, see Savulescu, 'Future People, Involuntary Medical Treatment in Pregnancy and the Duty of Easy Rescue'; Herring, *Medical Law and Ethics*, 169. I have discussed the justification of non-consensual medical interventions in criminal justice in Pugh and Douglas, 'Justifications for Non-Consensual Medical Intervention'.

⁶⁴ Eyal hints at something similar in Eyal, 'Paternalism, French Fries and the Weak-Willed Witness'.

⁶⁵ Mill, *On Liberty*, 81.

of such powers with regard to some claims is best justified by appealing to the strength of the interest that the claims protect.⁶⁶

Accordingly, even with regards to primarily self-regarding action, autonomy is not understood in the law to serve as a sufficient basis for the moral justification of any such action, as the consumerist view of autonomy would hold. In order to assess whether the law should delimit the scope of this individualistic aspect of the harm principle, we need to consider the relationship between beneficence and autonomy in more detail.

In the medical context, the principle of beneficence is sometimes understood in an extremely narrow sense to pertain only to medical benefits, that is, benefits concerning the end of healing.⁶⁷ On this interpretation, the goal of healing does not merely take precedence over other goods; it is the *only* good to which the principle of beneficence pertains. This interpretation is implausibly narrow; the reason for this is that we commonly use biomedical technologies in order to pursue ends that go beyond mere healing, and it can clearly be in our interests to do so. Although much depends on how widely we define the concept of health in our understanding of the narrow interpretation, vision enhancement surgery, or cosmetic procedures are examples of such interventions that can indisputably go beyond mere healing, but that are still plausibly in the individual's interests.

In view of this inadequacy, we need to refine the narrow interpretation of beneficence. Even if we maintain that medical interests should play a particularly salient role in a narrow understanding of the principle of beneficence in medical ethics, a plausible understanding of the principle should allow it to incorporate a broader range of benefits. As such, I suggest that we should understand the narrow interpretation of the principle of beneficence to be making the more plausible claim that medical benefits should override other kinds of benefit that might contribute to well-being. This allows for the possibility that interventions that do not aim at healing can still benefit the patient, whilst still affording a particularly salient role to health benefits in this conception of beneficence.

We can contrast these narrow interpretations of the principle, to a broad conception of beneficence, according to which the principle can be taken to pertain to *any* prudential benefits, without assigning any particular weight to a specific category of those benefits. To illustrate the difference, on the narrow interpretation of the scope of beneficence, it is difficult to see how death could ever be in a person's interests; such a choice is clearly contrary to the ends of healing, and this end takes precedence over other prudential goods on a narrow conception of beneficence. However, as I explained in my discussion of Isobel's case in the previous chapter, it seems possible for one to have a future life of prudential disvalue. In such cases, the broader conception might claim that death can be in a person's interests.

The extent to which the principles of beneficence and autonomy come into conflict depends significantly on our understanding of the scope of the principle of

⁶⁶ For instance, on similar matters, Foster writes: 'There will often be countervailing interests so powerful that they will outweigh autonomy interests. No one's autonomy right entitles them to be given poison, for instance'. Foster, *Choosing Life, Choosing Death*, 89.

⁶⁷ Pellegrino, *For the Patient's Good*.

beneficence. According to a commonly endorsed view that implicitly endorses something like the refined narrow conception that I illustrated above, the concepts of autonomy and beneficence are understood to represent two distinct domains; the question of what is in a patient's best interests is understood to be a conceptually different question to the question of what a patient autonomously desires. For instance, in introducing the concept of beneficence, Beauchamp and Childress point out that '[m]orality requires not only that we treat persons autonomously... but also that we contribute to their welfare'.⁶⁸

I shall argue that we should reject the narrow interpretation of beneficence that the commonly endorsed view relies upon. I have already suggested that we should understand the principle of beneficence to encompass goods beyond the ends of healing, and to pertain to prudential goods more broadly. In doing so though, we should also acknowledge the point that I have defended in this chapter, namely that autonomy plays an important role in a person's well-being. Thus, contrary to the commonly endorsed view, treating persons autonomously and contributing to their welfare should not be understood as distinct requirements; in order to adequately contribute to a person's welfare, we must take into account the agent's own autonomous preferences.

The problem with even the refined narrow conception is that it relies on an overly objective account of what is in a person's interests. Recall that on purely objective accounts of well-being, there are certain things that are intrinsically good or bad, that all agents have impersonal self-interested reasons to either want or avoid, regardless of their own attitudes towards these outcomes. The narrow interpretation of the principle of beneficence in medical ethics takes the end of healing to be the primary objective good of concern here. However, such a view is unattractive. This claim may seem somewhat surprising, since in section 1, I endorsed an explanatory account of well-being that appeals to objective values. However, what is problematic about the view that I am considering here is not that it relies upon the claim that there are objective elements of well-being. Rather, the problem with the view that I am considering here is that it implicitly assumes that there is an objective *ranking* of the different objective elements of well-being.

This assumption is problematic because the decisions that we make in medical contexts concern a far greater range of goods than those that are adequately captured by the end of healing. Our choices in this domain can have implications for pursuing the various other goods we may value in our life, and it is a mistake to assume that rationality requires that we must prioritize health over the promotion of these other goods. Indeed, whilst the interpretation of 'best interests' in medical law traditionally endorsed the narrow interpretation outlined above, this has shifted towards a broader conception of beneficence. For instance, best interests assessments under the MCA incorporate consideration of non-medical issues such as 'the person's past and present wishes and feelings' and also the 'beliefs and values that would be likely to influence his decision if he had capacity'.⁶⁹ This is an example of the way in which

⁶⁸ Beauchamp and Childress, *Principles of Biomedical Ethics*, 202.

⁶⁹ Mental Capacity Act 2005, 4(6). *Aintree University Hospitals NHS Foundation Trust v. James*, paragraph 24. Herring et al. argue that this interpretation of best interests is also echoed in the

the requirements of beneficence are not neatly separated from considerations of autonomy in medical law.

Even on theories of well-being that incorporate *only* objective elements, agents can still rationally disagree about the relative strengths of the self-interested reasons that different objective goods imply; I have suggested that truths concerning the comparative strength of such reasons are often very imprecise. Thus, even if we accept a purely objective list theory of well-being, we need not accept the claim implicit in the narrow interpretation of beneficence, that the goods in this list must have a set impartial degree of goodness, or that there is a supreme value that overrides others on the list. This point is all the more powerful if we endorse a hybrid view of the sort that I sketched at the end of section 1, which incorporates both objective and subjective elements of well-being.

On some objective views of well-being, such as that which is endorsed by the narrow interpretation of the principle of beneficence, conflicts between the principle of autonomy and the principle of beneficence will typically arise whenever an agent's autonomous desires conflict with the objective ranking of values that the view may stipulate. Such views of well-being naturally lend support to two types of paternalism; first, what Feinberg terms 'hard paternalism', and second, the sense of 'strong paternalism' that I explained above. According to hard paternalism, a third party may permissibly interfere with even an agent's voluntary choices in order to protect them from the harmful consequences of those choices; by way of contrast, soft paternalism only permits a third party to interfere with an agent's involuntary choices.⁷⁰ We may note that hard and strong paternalism are not necessarily co-extensive. For example, Sarah Conly's so-called 'coercive paternalism' is hard but weak, in so far as it allows the state to force people to act (or refrain from acting) in certain ways and impose actions on them that they would not choose (even if properly informed), but only in order to ensure that individuals are better able to achieve their own autonomously chosen goals.⁷¹ For instance, Conly argues in favour of banning cigarettes on the basis that doing so would be likely to advance individuals' effective pursuit of their long-term goal of better health.⁷²

The objection that I have raised against the narrow interpretation of beneficence, and the arguments that I have raised in favour of the final value of autonomy speak against both hard and strong forms of paternalism. First, contrary to strong paternalism, we seldom have warrant for assuming that rational agents should prioritize a particular goal, such that we would be warranted in determining that they pursue that end over other goods. The truths governing the relative weights of objective goods are imprecise, and rational agents can disagree about the relative weight they assign to the goods (such as health and pleasure) that are mutually incompatible in a particular context. Second, with respect to hard paternalism, if we believe that autonomy has final prudential value, then it may be a mistake to claim that overriding our autonomous choices is actually in our interests. Autonomy makes a

Montgomery judgement, which I considered in an earlier chapter. See Herring et al., 'Elbow Room for Best Practice?', 9.

⁷⁰ Feinberg, *The Moral Limits of the Criminal Law*, 12–13.

⁷¹ Conly, *Against Autonomy*, particularly 45. ⁷² *Ibid.*, 169–72.

particular kind of contribution to well-being that cannot be replicated by the imposition of other good things. This is precisely the insight of the PDA.

The claim that autonomy has final prudential value entails that there is a great deal of overlap between the values of autonomy and beneficence. This in turn lends support to the Harm Principle's contention that considerations of an individual's well-being cannot outweigh those of their autonomy. For this reason, many of the conflicts that are ostensibly conflicts between the two are more aptly construed as conflicts between different elements of well-being. This provides a philosophical basis for criticizing those elements of the law that appear to prioritize strong welfare interests over individual autonomy. Naturally, considerations of autonomy lend little support to claim rights that are not attended by the power of a waiver. More strikingly though, given the role of autonomy in well-being, we can also coherently challenge whether such rights are sufficiently supported by considerations of individual well-being. The claim that the law should safeguard unwaivable claim rights against *inter alia* sado-masochistic interactions or unusual body piercings, on the basis of the strength of the interests they protect, is to adopt an impoverished view of well-being that does not adequately capture the particular contribution that autonomy makes not only to a good life itself, but also to the nature of the contribution that other goods in that life can make to a person's well-being.

On the view that I am outlining, conflicts between autonomy and beneficence will be less commonplace; as long as an individual's choice is autonomous, that should give us at least a *pro tanto* reason to believe that respecting that choice will benefit that person, not because the choice is likely to lead to greater happiness (as the explanatory hedonist might claim), but rather because on this view there is prudential value to directing the course of one's life in accordance with one's own beliefs about what is of value, and with one's own beliefs about which values should take precedence.

However, the view that I endorse does not entail that there *cannot* be conflicts between autonomy and beneficence. Rather, I have suggested here that our analysis of cases in which there appears to be such a conflict should be more nuanced. Contrary to the overly objectivist account that the narrow view of beneficence implies, the fact that an agent has an autonomous preference is a fact that is relevant to our assessment of what is in their best interests. If an individual harbours an autonomous preference to engage in a behaviour that they believe they have reason to perform but that also endangers their life, such as the base-jumping thrill-seeker, or the gourmand who eats to the point of morbid obesity, there is a sense in which that behaviour is in their interests. However, whether that behaviour is in their *best* interests depends on the strength of their self-interested reasons to refrain from that behaviour, and the weight that we ascribe to autonomy in our general account of well-being.

One might worry that the principle of beneficence becomes superfluous on this approach, since it seems to have been subsumed by the principle of autonomy.⁷³ One

⁷³ Buchanan raises this sort of concern about this approach. See Buchanan, 'The Physician's Knowledge and the Patient's Best Interest', 94.

response to this worry is to interpret the principle of beneficence in a purely negative sense, by understanding it as ‘...an admonition to the physician not to allow the interests of others...to compromise his or her commitment to the patient’.⁷⁴ However, this reading seems to give the principle too narrow an interpretation, especially given that it may need to be operative in cases where we must consider the positive care of an individual who lacks autonomy.⁷⁵ Yet, the principle of beneficence can still have substance on the understanding of well-being and autonomy that I have outlined in this chapter, because it does not entail that the realization of autonomous choices *exhausts* the concept of well-being. The concept of ‘beneficence’ can incorporate both the patient’s autonomous choices, and other goods that agents have impersonal self-interested reasons to want. On this understanding, conflicts will arise when the agent’s autonomous choice is not co-extensive with what they have an impersonal self-interested reason to want. To resolve the conflict, we must compare the strength of these impersonal self-interested reasons, with the reasons we have to safeguard the agent’s autonomy by respecting their own assessment of what is in their best interests.

That the principle of beneficence still has substance on this understanding becomes further apparent when we consider cases in which a patient has not and cannot make their choices clear to their physician. For example, considering what agents have impersonal self-interested reasons to want can give physicians guidance when they are dealing with incompetent patients who lack a surrogate decision-maker or when asking for a competent patient’s consent would be too time-consuming in an emergency situation. Moreover, as Buchanan suggests, physicians need to have their own understanding of the patient’s best interests when deciding which courses of treatment to provide as viable choices for a particular patient and when making a recommendation to patients.⁷⁶

Conclusion

I have argued that whilst autonomy may often bear instrumental prudential value, it is primarily prudentially valuable for its own sake. I have also claimed that this implies that we should broaden our understanding of what is in an agent’s best interests. However, it should be acknowledged that I have not attempted to assign a particular definite value to autonomy. I have simply suggested that autonomy represents a special type of value for us, one that serves to amplify the contribution of other goods to well-being, in so far as there is a particular value in living a life that is our own. Yet, I have left open the possibility that this value can have different weight in different people’s conceptions of the good life. The extent to which autonomy will contribute to a person’s welfare will depend in part on the extent to which autonomy conflicts with other outcomes that the agent has reasons to pursue.

These claims are compatible with rejecting a consumerist view of autonomy that understands autonomy to be both necessary for and sufficient to the moral justification of medical interventions. I have argued that the view of autonomy’s value that

⁷⁴ Ibid., 95.

⁷⁵ Ibid.

⁷⁶ Ibid., 95–6.

I have presented here does not entail the consumerist view. At most, the view of autonomy's value that I have presented suggests that respect for autonomy, broadly conceived, is very often necessary to the justification of permissible medical interventions that are purely self-regarding. Given the central role that autonomy plays in well-being, narrowly construed beneficence-based justifications of non-consensual interventions are likely to come undone. For this reason, although the consumerist view should be rejected, autonomy still has a considerable bearing on our understanding of permissible self-regarding decisions in bioethical contexts.