

How Autonomy Can Legitimate Beneficial Coercion

Lucie White

Accepted Manuscript version. Published (2017) in

J. Gather, T. Henking, A. Nossek and J. Vollmann (eds.) *Beneficial Coercion in Psychiatry? Foundations and Challenges* (pp.85-99) Münster: Mentis.

Introduction

Respect for autonomy and beneficence are frequently regarded as the two essential principles of medical ethics, and the potential for these two principles to come into conflict is often emphasised as a fundamental problem (Beauchamp and Childress, 2001; Dunbar, 2003). On the one hand, we have the value of beneficence, the driving force of medicine, which demands that medical professionals act to protect or promote the wellbeing of patients or research subjects. On the other, we have a principle of respect for autonomy, which demands that we respect the self-regarding decisions of individuals. As well as routinely coming into opposition with the demands of beneficence in medicine, the principle of respect for autonomy in medical ethics is often seen as providing protection against beneficial coercion (i.e. paternalism) in medicine. The principle of respect for autonomy is achieving increasing precedence in medical ethics (O'Neill, 2002; Song, 2013), carving out a domain in which beneficence-based judgements should not operate. According to this dominant line of contemporary thought, we are required to respect the autonomous decisions of the individual; in cases in which the individual is making an autonomous decision in a medical context, it is inappropriate to ignore or override this decision on beneficence-based grounds. Conversely, paternalism, understood as overriding the autonomous, self-regarding decisions of the individual on beneficence-based grounds, is increasingly seen as taboo (Dunbar, 2003, p.39). However, these two values are not as straightforwardly opposed as they may appear on the surface. In fact, the way that we understand autonomy can lead us to implicitly sanction a great deal of paternalistic action, or can smuggle in paternalistic elements under the guise of respect for autonomy. This is because autonomous action is not simply akin to intentional action; in order for an action to qualify as autonomous, it must meet additional criteria. This feature of autonomy leads to a parallel between standards of respect for autonomy and beneficence in medicine; both may sanction ignoring or overriding certain of the patient's intentional, self-regarding decisions, because they do not qualify as autonomous, or because they are not in the patient's best interests, respectively.

This paper will be dedicated to outlining three ways in which the principle of respect for autonomy, depending on how we understand the concept of autonomy, can sanction or smuggle in paternalistic elements. As the specific relationship between respect for autonomy and beneficence will depend on how we conceive of autonomy, I will begin by outlining two dominant conceptions of autonomy, both of which have great influence in medical ethics. I will then turn to the three ways in which how we understand or employ autonomy can increase or support paternalism: firstly, when we equate respect for autonomy with respect for persons; secondly, when our judgements about what qualifies as an autonomous action contain intersubjective elements; and thirdly, when we expect autonomy to play an instrumental role, that is, when we expect people, when they are acting autonomously, to act in a way that promotes or protects their own wellbeing. I will then provide a proposal for how we might work to avoid this. I will suggest that it may be impossible to fully separate paternalistic elements out from judgements about autonomy. Instead, we are better off looking at why we are *motivated* to use judgements about autonomy as a means of restricting the actions of patients or research subjects. I will argue that this is a result of discomfort about speaking directly about our beneficent motivations in medical ethics. Perhaps we can reduce the incentive to smuggle in these beneficent motivations under the guise of autonomy by talking directly about beneficent motivations in medicine. This will also force us to recognise paternalistic motivations in medicine when they appear, and to justify paternalism where it occurs. I will bring this out with reference to two case studies.

Before I turn to these arguments, however, it is worth providing a brief note on my use of the word 'paternalism'. As we have already seen, some theorists suggest that paternalism only involves a restriction of autonomous choice – if the choice is not autonomous, then it is not paternalistic to override it. This is not the definition of paternalism that I will be working with here. Rather, I will take paternalism to involve any intentional interference in the self-regarding actions of another person (whether autonomous or not) with the aim of benefitting (or avoiding harm to) that person (see Beauchamp and Childress, 2001, p.178). According to this definition, any instance of beneficial coercion is an instance of paternalism.

The reasons for endorsing this definition of paternalism here are twofold. Firstly, a definition that suggests that it is only paternalistic to override autonomous action can provide dispensation to ignore or override decisions designated non-autonomous without acknowledgment or justification. This can license overriding a significant number of self-regarding decisions, as I will explore below. I wish, here, to make every instance of beneficial coercion apparent. Secondly, this

definition of paternalism is not pejorative – it provides no indication of whether paternalism is justifiable or not. Because respect for autonomy is now often regarded as of such fundamental importance, a definition of paternalism that is linked to autonomy will often have a negative connotation – to interfere with the autonomously chosen actions of another is seen as unacceptably disrespectful. I want to free paternalism from this connotation, in order to take seriously the possibility that paternalism may sometimes be justified, even when the individual is acting autonomously. I take paternalism, then, to be in need of recognition and justification, but not necessarily unjustifiable or wrong. As we will see below, a lack of prejudice against paternalism is crucial if we want to separate hidden beneficence-based motivations from standards of autonomy.

Two Theories of Autonomy

The distinction between intentional and autonomous action can be highlighted through reference to two cases much discussed in theoretical work concerning the concept; someone with obsessive-compulsive disorder compulsively washing his hands, and a drug addict succumbing to her desire for drugs. Though these actions are undertaken intentionally, there is something about them that appears to be problematic. Theories of autonomy often try to pinpoint what is problematic about these types of actions, generally by providing additional requirements beyond intentionality which specify what qualifies as an autonomous action. What form do these additional requirements take? Before we can interrogate the relationship between autonomy and beneficence, and how the concept of autonomy might be used to support beneficial coercion in some circumstances, we must have a working notion of what is meant by autonomy. However, pinning down this concept is a complicated task. Not only are there many different ideas about what autonomy amounts to in philosophy and medical ethics, but medical and research ethics guidelines tend to borrow elements from various ideas of autonomy, and tend to be vague about what notion of autonomy they are appealing to. Though this vagueness is worth keeping in mind, it does not mean that reference to specific models of autonomy will not be an important means of elucidating certain ways in which autonomy can legitimate beneficial coercion. I will draw these out with reference to two popular notions of autonomy.

The first theory I will refer to is the *authenticity-based* model of autonomy – an action counts as autonomous if it is *authentic*. There are many theories that fall into this camp, and the details differ, but there are two broad elements that unite these types of theories. Firstly, for an action or desire to qualify as autonomous, it must have some sort of special connection to the self. This relationship

can be construed in various ways: perhaps one needs to reflect upon and endorse one's desires in terms of the self, perhaps identification or coherence with the self (without explicit reflection) is required, or perhaps the desires need only not be repudiated on the basis of the self. Secondly, the 'self' that is referred to here should be understood as broadly Lockean; that is, as comprised of our enduring desires, values and other psychological elements (Taylor, 2005, p.17). On this conception, the drug addict and compulsive hand-washer's desires may fail to qualify as autonomous due to a failure to meet the condition of authenticity. The agents here may, upon reflection on the basis of their enduring desires and values, reject these desires, or these actions may not fit with these desires or values (whether they reflect upon them or not). Or, if the authenticity condition is construed at its weakest, perhaps the drug addict and hand-washer repudiate their actions; although they undertake them intentionally they see them as alien to themselves, as actions that they are driven to undertake against their better judgement. This notion of autonomy is prominent in theoretical philosophical work, though its influence extends to medical contexts. Some aspects of this view, as we will see, have been particularly influential; the understanding of the significance of autonomy and what the purpose of a principle of respect for autonomy is in medical ethics have been shaped, in part, by the philosophical tradition leading up to this notion.

The second model, which I will refer to as *practical* autonomy, suggests that we should not focus on authenticity when we are talking about which actions are autonomous. Rather, actions should count as autonomous if they are undertaken intentionally, with a sufficient level of understanding, and are sufficiently free from controlling influences (see Faden and Beauchamp, 1986; Beauchamp and Childress, 2001). The problem with the drug addict and compulsive hand-washer's behaviour, according to this model, would likely be traced to a failure to act sufficiently free from controlling influences, understood as including internally controlling influences such as those arising from compulsion or addiction. This approach to which decisions count as autonomous, developed by bioethicists Ruth Faden, Tom Beauchamp and James Childress, has been enormously influential in bioethics, and its influence can be seen in some informed consent guidelines, designed to protect autonomy in practical contexts (see, for example, The National Commission, 1979).¹ While the authenticity-based model of autonomy has taken theoretical metaphysical concerns as a primary focus, this model has been developed for use in practical situations. It is particularly well-suited to this task for two reasons; firstly, these three conditions capture the wide range of actions that we

¹ For more on the connection between this document and the approach in the works referred to above, see Beauchamp, 2005.

typically think of as worthy of respect, and secondly, compared to a condition of authenticity, these requirements are easier to assess.

How Autonomy Can Legitimate Beneficial Coercion

1) Respect for autonomy is seen as akin to respect for persons.

The first way in which our judgements about autonomy can sanction paternalism, or smuggle in paternalistic elements, is when respect for autonomy is equated with ‘respect for persons’. This is brought out most clearly with reference to the authenticity-based model of autonomy, because the idea that respect for autonomy is a natural expression of respect for persons makes a lot of sense in this context. However, as we will see, this is *not* exclusive to those who subscribe to an authenticity-based view of autonomy.

As we have seen, authenticity-based autonomy requires that an action is *authentic* in order to qualify as autonomous. More specifically, on the basis of standing in some special relation to the enduring desires and values that constitute the self, the action counts as authentic – it has a special link to the self. This is a useful feature of autonomy, because it provides us with a way, in practical contexts, to express the abstract value of ‘respect for persons’. That is, if we want to show respect for persons, we should show respect for the actions that form a special expression of the self. There is, however, a distinct disadvantage to this useful aspect of autonomy. If we equate the value of respect for persons with respect for their autonomous desires and actions, this means that we can ignore any desired courses of action that do not qualify as autonomous without violating the principle of respect for persons. We do not have to give any further, paternalism-based justification for overriding these sorts of desires.² However, this can sanction ignoring intentional, self-regarding decisions; designating them as unworthy of respect and, therefore, as free to be overridden on beneficence-based grounds, all without invoking an explicitly paternalistic justification. In a medical context, where healthcare professionals are motivated by a concern for the wellbeing of the patient, we can expect that any decisions which are not protected under a principle of respect for autonomy, or not seen as worthy of respect, will be overridden when they are not seen as furthering this goal.

This will be a particular problem when the standards for what qualifies as an autonomous decision are quite stringent; that is, when they designate many of our self-regarding decisions as non-

² This dovetails with the alternate view of paternalism discussed above; that paternalism only involves overriding autonomous actions, and that it is inherently disrespectful – if overriding an action does not violate ‘respect for persons’, according to this line of reasoning, it is not disrespectful, and thus does not constitute paternalism.

autonomous and therefore free to be overridden. Authenticity-based standards for autonomous action, in addition to providing a link between the decisions designated as autonomous and the value of respect for persons, can be relatively demanding. In other words, authenticity-based autonomy tends to pick out a relatively small subset of our desires and actions as autonomous, as well as providing a basis for ignoring or overriding the rest.³ Critiques of this notion of autonomy have tended to focus on the stringency of the standards, rather than the claim that respect for autonomy can be equated with respect for persons. This concern is the motivation behind the development of what I have referred to as practical autonomy, which rejects an authenticity condition as a necessary criterion of autonomous action, focusing instead on the more permissive and intuitive conditions of intentionality, understanding and noncontrol. However, the idea that non-autonomous decisions may be overridden or ignored without violating the principle of respect for persons, and without appealing to paternalism, is still retained in this model. In addition, these new conditions for autonomy bring new problems to the fore.

2) *Intersubjective elements in autonomy.*

By introducing the conditions of understanding and noncontrol into the equation, the practical model introduces *intersubjective elements* into our judgements about autonomy. The condition of understanding provides an illustrative example here: how do we judge whether someone sufficiently understands a proposed course of action? Imagine a marathon runner who has sustained damage to his leg and can no longer run. He wishes to undergo a new and experimental surgery which has a 10% chance of repairing his leg, but a 50% chance of confining him to a wheelchair. This person might value running so much – it might be of such central importance to his life – that he is willing to undergo this substantial risk in order to restore his ability to run. But someone assessing whether the runner understands the risks of the decision might find this decision crazy or incomprehensible. Consequently, he might judge that the runner cannot possibly have an adequate understanding of the consequences of his decision.

Though authenticity-based autonomy is restrictive, its great advantage lies in the fact that it gives us a *procedure* for determining autonomy without imposing our own values onto the person. What is important for authenticity-based autonomy is that a person's desires match up with their higher-order, self-forming desires, but the *content* of these desires is out of reach of scrutiny. This does not mean that there is no possibility for our intersubjective judgements to creep in under an

³ For a comprehensive overview of the various ways in which the authenticity condition can be construed, and how each of these formulations result in an overly restrictive or otherwise inappropriate criterion for a sufficiently permissive and intuitive account of autonomous action, see Faden and Beauchamp, 1986, pp.262-268.

authenticity-based view of autonomy: we might, on the basis of our distaste for a certain course of action, assume that this must be due to a lack of authenticity, and thus judge the proposed action non-autonomous on these grounds. But this problem is certainly going to be more significant when we abandon a procedural framework, theoretically independent from judgements about the content of a person's desires or decisions. In cases in which we must evaluate whether someone adequately understands a proposed course of action, or whether they are being driven to a decision by controlling influences, it becomes very difficult to avoid introducing intersubjective judgments into the equation. Someone else will need to judge whether a person meets the requirements of understanding and noncontrol for their decisions to be classified as autonomous, and there is no framework for assessing these variables which is independent of the judgements of others. There is thus increased potential, under this understanding of autonomy, for paternalistic elements to sneak into our judgements about whether someone is acting autonomously.⁴

3) *Autonomy expected to play an instrumental role.*

This brings us to the third problem: on what grounds do we judge that someone displays sufficient understanding, or is sufficiently free of controlling influences, to qualify as autonomous? What is likely to play a role here is that we generally expect people to act in a way that promotes and protects their own interests. People generally have a strong motivation to act in this way, and they have knowledge about their own tastes and values that can put them in a good position to pursue their own happiness.⁵ Autonomy, in other words, can have significant instrumental value. Because we are likely to value making our own decisions (in part) for instrumental reasons (Scanlon, 1998), we have a tendency to expect that people will use their autonomous decisions as an instrument to pursue their wellbeing. When we do not judge that they are doing this, we are likely to suspect that there must be a problem with autonomy.

There are two reasons, however, that someone's autonomously chosen course of action may not correspond with our conception of what is in their best interests.⁶ Firstly, someone may

⁴ It is worth reiterating here that, just as there is still potential for intersubjective judgements to colour a judgement about whether an action is authentic, the danger that autonomy is used as a means of overriding decisions on the basis that they do not meet the standards of autonomy (rather than invoking an explicitly beneficence-based justification) is retained in the practical model of autonomy. Although the criteria here are more permissive, and will thus exclude fewer decisions on these grounds, this model of autonomy does not abandon the idea that respect for autonomy is an appropriate means of showing respect to persons.

⁵ This view is most famously expounded in J. S. Mill's *On Liberty* (1986). I certainly do not wish to maintain, as Mill does, that freedom of choice always has these fortuitous effects (as will become apparent below), only that it can, and that this can form an expectation on our part that people will act in this way (if they are acting freely or autonomously).

⁶ In addition, of course, to simply making a miscalculation about the consequences of one's actions.

autonomously choose something that they themselves do not conceive of as being in their own best interests. Self-sacrificing behaviour or acting according to our moral values can meet the criteria of both standards of autonomy outlined above, but may not correspond to what a person, even subjectively, feels is in her own best interests (White, 2013). Secondly, there can be a gap between what someone subjectively judges to be in her own best interests (subjective wellbeing), and what another may judge to be in her best interests (intersubjective wellbeing). The runner case above provides an example of this. The runner may be choosing what, on the basis of his own tastes and values, he sees as the best course of action in terms of the pursuit of his own wellbeing. A third party, however, may see the assumption of so much risk as clearly against the best interests of the runner. Coupled with the belief that people act, when they act autonomously, to pursue their own wellbeing (even imagining an exception for acting on moral principles, as outlined above), the third party must conclude that the proposed course of action results from a failure of autonomy (perhaps stemming from a failure to adequately understand the likelihood of the various possible outcomes). The second and third elements outlined here, then, work closely in tandem with each other. Though there is more scope for this kind of assumption to colour our judgements in the practical model of autonomy, this third aspect is not tied so closely to a specific model of autonomy. Rather, this stems from the reasons we value autonomy, and the role that we expect autonomous decision-making to play in a medical context.

What can we do about this?

I have outlined three ways in which autonomy can sanction beneficial coercion or smuggle in paternalistic elements. Firstly, when we equate respect for autonomy with respect for persons, thus allowing any decisions that do not qualify as autonomous to be overridden without violating the principle of respect for persons, and without explicitly paternalistic justification. Secondly, when our judgements about what counts as an autonomous action contain intersubjective elements, especially when we are evaluating whether someone sufficiently understands a course of action, or whether they are operating sufficiently free of controlling influences. And thirdly, closely tied in with the second point, when we expect people to use their autonomous decisions to pursue their own wellbeing – when we do not see how they are doing this, we are less likely to judge that they are acting autonomously.

In the remainder of this paper, I will turn to the question of what might be done about this. I contend that it is impossible to completely remove the potential for these three elements to enter into our judgements about which actions count as autonomous. Judgements about autonomy will

always involve excluding some self-regarding actions, and on good grounds; where actions are undertaken in ignorance of the consequences, where they are coerced, or where they are the result of compulsion or addiction, there may be good reason to refuse to honour and endorse them. In addition, *we* will always make judgements about whether a given action is autonomous – judgements about autonomy are always taking place in a wider intersubjective context. It is an inescapable feature of medicine that these judgements will always be made intersubjectively and, as long as that is the case, it is difficult to remove the possibility that those who must make these judgments will tend to classify actions that seem bizarre or incomprehensible as lacking in autonomy.

The Case for Beneficence

Thus, rather than trying to remove the potential for paternalistic elements to sneak into our judgments about autonomy, I suggest we take a slightly different approach. This will involve, rather than focusing purely on *how* paternalistic elements can be incorporated into judgements about autonomy, asking *why*. Why might we want to sneak paternalistic elements in to our judgements about medical decisions under the guise of autonomy? What might be the motivation for using autonomy in this way? The answer can be found in the considerations with which I opened the chapter, summarized here by Terri Dunbar (2003, p.39):

Contemporary opinion suggests that where there is tension between autonomy and beneficence, autonomy should take precedence...[Actions] which override the patient's autonomy are considered paternalistic, a construct which has achieved almost taboo status.

We see autonomy as having overriding force; we are reluctant to entertain the idea that autonomy, where it exists, should be overridden. On the same note, we are reluctant to think that paternalism (where paternalism is more narrowly construed as interference with only autonomous action) can ever be justified. This puts pressure on autonomy to serve as a means for sorting the desires that we do not wish to honour and respect from those that we do. If we achieve this under the guise of autonomy, we do not have to invoke a paternalistic justification for overriding these desires; we can simply say that they do not qualify as autonomous. I suggest, therefore, that we might remove some of the motivation for using autonomy as a means of imposing paternalistic restrictions by acknowledging our beneficence-based motivations where they occur, and considering the idea that acting from beneficence-based reasons may sometimes be justified, *even where* the individual is

acting autonomously. This will also force us to explicitly acknowledge the beneficence-based restriction of actions where it occurs. Where we are acting paternalistically, we should recognise and justify this.

This is best brought out through the examination of specific case studies. I will now turn to two situations in which, I will argue, we have an intuitive desire to prevent certain behaviour. In these cases, as we will see, attempts are made to justify this by suggesting that this behaviour is not autonomous. Perhaps, however, we should acknowledge that our concerns here are more likely motivated by considerations of wellbeing.

Case 1: Bial Drug trial

The recent Bial drug trial in France, or more specifically, the media reaction to this event, displays this use of autonomy particularly clearly. In January 2016, five participants experienced serious adverse reactions while taking part in a Phase 1 drug trial (where a drug is provided to healthy volunteers in order to test its safety) in France. One man lapsed into a coma and was declared brain-dead, while five others were hospitalized, with four displaying symptoms of brain damage (TSSC, 2016). In surveying media coverage of this event, a pronounced trend becomes apparent: attempts to pinpoint what went wrong with this trial focus heavily on issues to do with autonomy and consent. This took the form of three main streams of criticism. The first focused on the idea that there must have been ‘undue inducement’ in this case. That is, the payments for participating were so high that they functioned as a *controlling influence*, forcing the men to participate. This can be displayed through the contention that volunteers for Phase 1 trials are typically “poor and unemployed, and they volunteer for trials like this because they are desperate for money” (Chan, 2016). Participants are thus “easily exploited” (Chan, 2016) or the trial involves “inherent coercion” (Stone, 2016). A second stream of criticism suggested that there must be a problem with the information that the subjects received; that they could not have possibly had an adequate *understanding* of the risks. The subjects “may not have been given all the info” (Tasker, 2016), or the investigators or consent form must not have adequately explained the possible risks (Stone, 2016). The third type of criticism shows instrumental assumptions coming into the equation through suggestions that the *very fact* that healthy volunteers consented to research with this level of risk shows that there was a problem with autonomy; a healthy volunteer simply would not autonomously consent under these circumstances (Stone, 2016).

Despite this focus, and although undue inducement and lack of information certainly pose problems with this type of research, they do not get to the core problem in this case. We can bring this out by asking, even if the subjects did autonomously consent to the procedure; even if they fully understood what they were doing and there was no undue inducement, would it be acceptable to subject these willing, autonomous volunteers to this level of risk? There is nothing to suggest that a subject *cannot* autonomously consent to this procedure. A research subject may choose to take the gamble, knowing the chance of serious side effects will be small, and decide that he could use the money (even though he does not feel compelled to take it). Even if this were the case, there is certainly a case to be made that he should not be able to participate anyway – that we, as a society, do not want to put people’s wellbeing at risk in this way even if they desperately and autonomously want to participate. We have a notion in research that there is ‘no injury where the subject is willing’, but this is not necessarily true. People can autonomously take risks, and they can autonomously behave in destructive ways. If we hold otherwise, we are inviting paternalistic controls around the judgement of who is autonomous. There might be good beneficence-based grounds to limit the amount of risk that we let even autonomous research subjects take on, but we should recognise that our concerns here are beneficence-based, and not try to limit this sort of research by attempting to find a problem with the autonomous consent of the research subjects.⁷

Case 2: Body Identity Integrity Disorder

The tendency to focus solely on autonomy and to exclude unwanted conduct on autonomy-based grounds is not exclusive to medical research, as we can see by turning to the treatment of body identity integrity disorder. This is a very rare condition in which a person feels that one of their limbs does not belong to their body, and wants to have it removed. In evaluating whether this desire should be honoured by physicians, most bioethicists focus heavily on autonomy as the deciding consideration (Song, 2013, p.491). The unquestioned acceptance that the ethical acceptability of this procedure hinges on the question of whether this desire is autonomous leads people, as in the case above, to attempt to show why this desire does not qualify as autonomous. Sabine Müller’s treatment of this issue provides an example of this. Based on the contention that “the crucial question is whether the amputation desire is an autonomous decision” (2009, p.42)

⁷ I should note here that I do not wish to rule out the possibility that, upon consideration, we might decide that autonomy is of sufficient importance to override our beneficence-based concerns (both in this case and in others, including the case below). I phrase these considerations in this manner only because we must take the possibility that paternalism might be justified seriously, in order to reduce the pressure on autonomy to play the role of sorting what we see as acceptable from unacceptable desires. We must be free to examine these considerations in terms of beneficence without the stigma that generally accompanies the thought of paternalism, and without the assumption that paternalistic concern cannot override autonomous action.

Müller argues, using both an authenticity-based conception of autonomy and by looking at the conditions of understanding and noncontrol from practical autonomy, that the desire to have a limb removed is not autonomous. Her treatment of this issue is perplexing, as these cases are particularly vexing for medical ethicists precisely because this desire *does* seem to be autonomous. This desire appears to be highly authentic; people who have these desires have typically had them since childhood, and they talk about them in terms of identity – they often say that they do not feel like their unwanted limb is part of themselves, or they experience it as alien. Furthermore, they are not operating under a delusion which might compromise *understanding* of the situation (Bayne and Levy, 2005; White, 2014). It is also dubious to suggest that the desire arises from an internally *controlling influence*. Unlike the obsessive hand-washer referred to above, this desire is not experienced as alien or as controlling the person against her will. There is no specific feature of this desire that would allow us to clearly distinguish it from any other persistent desire (Slatman and Widdershoven, 2009, p.49), unless we refer to the fact that this desire is symptomatic of a mental illness. There is no reason, however, to think that every desire that comes from mental illness should be automatically excluded from counting as autonomous. This is apparent when we look at gender identity disorder, to which body identity integrity disorder is frequently compared (see White, 2014). This is classified as a mental illness, but we do not see it as affecting autonomy. In fact, in this case, we see surgery as a means to increasing autonomy – to changing the body to match with the person’s authentic conception of the self.

What then, motivates this drive to show that this desire is not autonomous? I contend that this is a result of the fact that this desire just seems incomprehensible, even upsetting, to us, coupled with the idea that autonomy will be the crucial deciding factor in whether this desire should be brought to fruition. If the decision is indeed autonomous, we do not want to override it, and this leads us to attempt to establish that the decision is not autonomous. However, this focus on autonomy is again obscuring a more crucial question about beneficence. We can bring this out through the following question: what might be the difference between surgery to remove a limb and sex reassignment surgery? The difference here is not likely to be a difference in autonomy, but is, in fact, a question of harm. Should we conceive of this surgery as doing harm to the patient? Or should we conceive of it as alleviating the psychological suffering of the patient? If this was clearly doing harm to the patient, autonomous consent might not suffice – there might be good reason to say that beneficent concern for the patient justifies a restriction. Conversely, if this turns out to be in the patient’s best interests, this strengthens the case for it significantly. I do not wish to take a specific position on this issue here, rather, I want to suggest that our concerns about this case

are more fruitfully discussed in terms of beneficence. When we take for granted the idea that autonomy is the crucial deciding factor in this case, we obscure the beneficence-based issues underneath. The contention that autonomy is pivotal also leads to us using autonomy as a means of separating the desires that we find problematic from the ones that we see as acceptable and understandable. To open up the discussion about beneficence, and to prevent autonomy from being used in this way, we might have to accept that autonomous consent might not be sufficient to justify surgery in these cases. Paternalism may be justified here, but if it is paternalism, we should make sure that we are recognising what we are doing, and justifying it.

Conclusion

I have argued that there are three main ways in which autonomy can sanction beneficial coercion, or smuggle in paternalistic elements. By equating respect for autonomy with respect for persons, we can disregard or override certain desires without appealing to paternalism. Due to the presence of intersubjective elements in our judgements about who is autonomous, there is space for us to judge desires as non-autonomous if we do not understand or agree with them. And this is coloured by our expectation that people will use their autonomy as an instrument for pursuing their wellbeing, which makes us less likely to judge that a decision is autonomous if we do not see how it does this. Rather than trying to remove the possibility of autonomy being used in this way, I have suggested that we look instead at *why* it is used in this way. I have argued that this comes from a reluctance to talk about beneficence directly. I have therefore suggested that we entertain the idea that beneficent considerations play a role in evaluating the ethical acceptability of procedures, even where the subject is autonomous. As well as reducing the pressure on autonomy to smuggle in paternalistic elements, this forces paternalism out into the open, so we are forced to justify it.⁸

References

Bayne, T. and Levy, N., 2005. Amputees by choice: body identity integrity disorder and the ethics of amputation. *Journal of Applied Philosophy*, 22(1), pp.75-86.

Beauchamp, T., 2005. The origins and evolution of the Belmont Report. In: J. Childress, E. Meslin and H. Shapiro, eds. *Belmont revisited: ethical principles for research with human subjects*. Washington D.C.: Georgetown University Press. pp.12-26.

⁸ Thanks to Dietmar Hübner, Gregor Scherzinger and Thomas Schramme for their useful comments on an earlier draft of this paper.

Beauchamp, T. and Childress, J., 2001. *Principles of biomedical ethics*. 5th ed. New York: Oxford University Press.

Chan, S., 2016. 6 Hospitalized, one of them brain-dead, after drug trial in France. *The New York Times*, [online] 15 Jan. Available at: <http://www.nytimes.com/2016/01/16/world/europe/french-drug-trial-hospitalization.html> [Accessed 26 April 2016].

Dunbar, T., 2003. Autonomy versus beneficence: an ethical dilemma. *Primary Health Care*, 13(1), pp.38-41.

Faden, R. and Beauchamp, T., 1986. *A history and theory of informed consent*. Oxford: Oxford University Press.

Mill, J. S., 1986. *On liberty*. New York: Prometheus Books.

Müller, S., 2009. Body integrity identity disorder (BIID) – is the amputation of healthy limbs ethically justified? *The American Journal of Bioethics*, 9(1), pp.36-43.

O'Neill, O., 2002. *Autonomy and trust in bioethics*. Cambridge: Cambridge University Press.

Scanlon, T. M., 1998. *What we owe to each other*. Cambridge: Harvard University Press.

Slatman, J. and Widdershoven, G., 2009. Being whole after amputation. *American Journal of Bioethics – Neuroscience*, 9(1), pp.48-49

Song, R., 2013. Body identity integrity disorder and the ethics of mutilation. *Studies in Christian Ethics*, 26(4), pp.487-503.

Stone, J., 2016. Bial's clinical trial in France ends in disaster. What went wrong? *Forbes*, [online] 16 Jan. Available at: <http://www.forbes.com/sites/judystone/2016/01/16/bials-french-clinical-trial-ends-in-disaster-what-went-wrong> [Accessed 26 April 2016].

Tasker, R., 2016. Students share why they take part in clinical trials. *The Daily Collegian*, [online] 12 Feb. Available at: <http://www.collegian.psu.edu/news/campus/article_4b9527b2-d10a-11e5-9840-0f27804abe19.html> [Accessed 26 April 2016].

Taylor, J. ed., 2005. *Personal autonomy: new essays on personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press.

Temporary Specialist Scientific Committee (TSSC), 2016. Version anglaise: Minutes of the Temporary Specialist Scientific Committee (TSSC) meeting on “FAAH (Fatty Acid Amide Hydrolase) Inhibitors” of 15 February 2016. *Agence Nationale de Sécurité du Médicament et des Produits de Santé*, [online] Available at: <http://ansm.sante.fr/var/ansm_site/storage/original/application/5a8a7343ad970487b473274c67ab2db1.pdf> [Accessed 26 April 2016].

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (The National Commission), 1979. *The Belmont Report: ethical principles and guidelines for the protection of human subjects of research*. Washington: U.S. Government Printing Office.

White, A., 2014. Body identity integrity disorder beyond amputation: consent and liberty. *HEC Forum*, 26(3), pp.225-236.

White, L., 2013. Understanding the relationship between autonomy and informed consent: a response to Taylor. *Journal of Value Inquiry*, 47(4), pp.483-491.