The Need for Authenticity-Based Autonomy in Medical Ethics

Lucie White

Pre-print version. Published in *HEC Forum* (2017).

Available at Springer via **https://doi.org/10.1007/s10730-017-9335-2**

Abstract The notion of respect for autonomy dominates bioethical discussion, though what qualifies precisely as autonomous action is notoriously elusive. In recent decades, the notion of autonomy in medical contexts has often been defined in opposition to the notion of autonomy favoured by theoretical philosophers. Where many contemporary theoretical accounts of autonomy place emphasis on a condition of "authenticity"; the special relation a desire must have to the self, bioethicists often regard such a focus as irrelevant to the concerns of medical ethics, and too stringent for use in practical contexts. I argue, however, that the very condition of authenticity that forms a focus in theoretical philosophy is also essential to autonomy and competence in medical ethics. After tracing the contours of contemporary authenticity-based theories of autonomy, I consider and respond to objections against the incorporation of a notion of authenticity into accounts of autonomy designed for use in medical contexts. By looking at the typical problems that arise when making judgments concerning autonomy or competence in a medical setting, I reveal the need for a condition of authenticity—as a means of protecting choices, particularly high-stakes choices, from being restricted or overridden on the basis of intersubjective disagreement. I then turn to the treatment of false and contestable beliefs, arguing that it is only through reference to authenticity that we can make important distinctions in this domain. Finally, I consider a potential problem with my proposed approach; its ability to deal with anorexic and depressive desires.

Keywords Authenticity; Autonomy; Competence; Medical Ethics; Locke; Anorexia; Depression

Introduction

Though the concept of autonomy is central to both theoretical philosophy and medical ethics, the role of the concept in these two fields, and thus the development models of autonomy and the understanding of what qualifies as an autonomous desire or action, have diverged in fundamental ways. Theoretical accounts of autonomy have generally focused on the conditions under which a desire can be regarded as one's own, which has led to many accounts making a condition of "authenticity" a central requirement of autonomous action or desire. Accounts of autonomy and

competence designed for deployment in medical settings, in contrast, focus on conditions under which a proposed course of action should be respected, or not overridden. This has led to claims that these two concepts of autonomy have nothing to do with each other (Swindell 2009), and that authenticity should not be regarded as a condition of autonomy in medical contexts (Faden and Beauchamp 1986). I will argue, on the contrary, that authenticity should be regarded as playing a vital role in adequate models of autonomy and competence in medical settings. More specifically, authentic values form an essential underlying basis upon which the conditions of autonomy in medical settings can be evaluated. After sketching the outlines of theoretical authenticity-based autonomy, and considering the reasons for its exclusion in medical accounts, I will draw out the role of and need for reference to authentic values in medical accounts of autonomy. Namely, it is only by using authentic values as a frame of reference that we can make important distinctions between different types of false or contestable beliefs, and protect desires from being overridden due to intersubjective disagreements about their content. Finally, I will turn to some types of desire that seem to pose a problem for authenticity-based accounts of autonomy in a medical context, providing suggestions for how we might deal with them.

1. Authenticity-based autonomy

We can best trace the contours of contemporary, authenticity-based accounts of autonomy by starting with their predecessor; hierarchical accounts of autonomy. This type of account was developed independently by several theorists in the early 1970s, motivating a subsequent focus on "individualistic" conceptions of autonomy that are centered not on a direct restriction of the content of desires and actions, but rather their structure and relation to one another (Taylor 2005). This kind of approach is exemplified by Harry Frankfurt's extremely influential theory of personhood.[1] Aiming to devise an account of personhood that could adequately differentiate between persons and non-persons (such as animals), Frankfurt highlights the "structure of a person's will" (1971, p.6) as a distinguishing factor of persons. We cannot reveal the structure of a person's will by simply looking at their desires to undertake a certain action (first-order desires), he contends – what is characteristic of persons is their ability to have desires about their desires (or second-order desires). Frankfurt singles out the possession a particular type of second-order desire as a necessary condition of being a person, what he refers to as a "second-order volition"; a second-order desire to make a first-order desire "effective" (to translate it into action). If we can

[1] Frankfurt doesn't use the term autonomy here, but his theory as outlined here is often referred to as a theory of autonomy (see Taylor 2005), and when he began later to explicitly write about the notion of autonomy, he utilizes the same hierarchical account (Bratman 2007).

make the first-order desires that move us to action correspond[2] with our second-order volitions, Frankfurt contends, then we exercise what he refers to as "freedom of the will" (1971, p.15) – for our purposes, we can say that the person is autonomous with respect to this action and the desire to undertake it (Taylor 2005).

In order to determine whether an action is autonomous, then, we must, according to hierarchical accounts of autonomy, look at the hierarchy of desires concerning the choice, to determine whether they are in line with each other. This broad approach to autonomy has several attractive features. One, perhaps of most interest to theoretical philosophers, is that this notion of autonomy is compatible with physical determinism. On a related note, this notion of autonomy does not require a "self-created self" as a precondition of autonomous action – it is compatible with the fact that we are born into the world with given tastes and proclivities and develop our character and preferences within a given environment, and thus cannot create our characters from scratch. An additional desirable feature of hierarchical accounts of autonomy, as noted above, is that they are value neutral; any desire can count as autonomous, regardless of content, as long as it is supported and endorsed by an appropriate second-order volition. This provides us with a concept of autonomy which is useful in many contexts in applied ethics, where autonomy is often deployed as a means of creating a space in which people are able to pursue their own values, given the contestable nature of values in a pluralistic society (Taylor 2005).

However, hierarchical theories of autonomy are also beset by many problems, among them, a criticism that has come to be known as the "problem of authority". Critics of Frankfurtian-style hierarchical accounts of autonomy questioned what it is about second-order desires that confer on them the authority to designate other desires autonomous. Frankfurt notes that the ability to have these second-order desires is characteristic of persons, but this does not seem to suffice to give second-order desires this authority. Related to this is the "problem of regress" – why should we stop at second-order desires instead of evaluating these in terms of third-order desires, and so on? As a response to these and other concerns,[3] theorists began to develop increasingly sophisticated models of autonomy that drew on the central insights of hierarchical theories, but were constructed in ways which allowed them to avoid these problems. A central element in many of these theories is the introduction of a notion of the self, construed in broadly Lockean terms; that is, the self is

[2] Hierarchical theories of autonomy are sometimes thought to require explicit reflection on the basis of second-order desires, but this is something proponents of hierarchical theories sought to avoid; see Frankfurt (1971) and Dworkin (1988). I will thus focus on correspondence rather than reflective acceptance in what follows.

[3] For an account of the various problems faced by hierarchical theories of autonomy, and the various ways that subsequent theories of autonomy have been developed to avoid these problems, see Taylor (2005).

seen as a set of enduring, stable overlapping psychological elements, including values, beliefs and desires. This notion of the self can, roughly speaking, be seen as taking the place of second-order volitions in these models of autonomy – rather than a desire to undertake a certain action being autonomous in virtue of correspondence with a second-order volition, desires are autonomous in virtue of reflecting, corresponding to or cohering with the enduring desires, values and beliefs that constitute the self.[4] It is the requirement that a desire stand in some special relation to the (Lockean) self that comprises the *condition of authenticity*, and forms the foundation for *authenticity-based autonomy*.

Bringing this notion of the self into theories of autonomy avoids the two problems with hierarchical autonomy mentioned above. The problem of authority is avoided by replacing second-order desires with the self – if we take autonomy to inhere in authenticity, that is, the condition that autonomous desires or actions stand in a special relationship with the self, the self does indeed have the authority to confer autonomy upon desires and actions. The problem of regress is similarly circumvented, because the condition of authenticity just aims to gauge the relationship between any given desire or action and the enduring desires and actions that make up the self. Because these enduring desires and values *constitute* the self, we do not need to (in fact, it makes no sense to) ask whether these psychological elements are authentic to the self. (Noggle 2005). An authenticity-based notion of autonomy retains the advantages of value neutrality, and not requiring a self-created self. Introducing a condition of authenticity into theories of autonomy also involves additional advantages. Rather than simply pointing out an ability that is specific to persons, authenticity-based autonomy allows us to identify the desires and actions that form a special expression of the self. In respecting, protecting and facilitating these desires and actions, therefore, we are given a concrete means of honoring the abstract value of "respect for persons".

2. Faden and Beauchamp's criticisms of authenticity-based autonomy

Where theoretical accounts of autonomy are generally concerned with identifying the conditions under which one's actions or desires can be called one's own, the notion of autonomy in medical ethics takes a different focus. The concept of autonomy (and related concept of competence) in medical ethics are aimed towards discerning which decisions patients and research subjects should be able to make in a medical setting. This has led to calls to exclude authenticity as a condition of autonomy in these contexts (Faden and Beauchamp 1986) and to claims that philosophical notions

[4] For examples of the ways that this type of theory of autonomy actually incorporate the Lockean self, see Bratman (2007) and Ekstrom (2005).

of autonomy and concepts of autonomy and competence in medical ethics have nothing do to with each other, that it is inappropriate to use philosophical notions of autonomy to evaluate medical decisions, and that it is too difficult to assess whether a decision is autonomous in the sense required by theoretical philosophers (Swindell 2009). In their influential and practically-focused account of autonomy, Ruth Faden and Tom Beauchamp aim to produce a model of autonomy which will capture the decisions that we intuitively feel are worthy of the special protection that being designated autonomous confers. They suggest that actions[5] should be considered autonomous if they are *intentional*, substantially *understood*, and substantially *noncontrolled*. Due to the ubiquity of authenticity-based notions of autonomy, they consider, then reject, the idea of adding a fourth condition of authenticity.

They are concerned about a condition of authenticity for two reasons. First, they want their model of autonomy to present criteria for what it is for an individual *action* to qualify as autonomous. They point out that in authenticity-based autonomy, autonomy is seen as involving "a *kind* of agent", namely, the kind of agent whose "life has a consistency that derives from a coherent set of beliefs, values and principles, by which his actions are governed…[that] he has made his own." (1986, p.236). They suggest that this idea of autonomy is more focused on autonomous *persons*. The only way in which it is possible to evaluate whether one's decision is autonomous under a Lockean authenticity-based theory, for example, is with reference to the stable, enduring values that constitute the self. Faden and Beauchamp are concerned that this leads to a focus on whether agents possess the capacity for autonomy (i.e. whether they have the kind of stable or coherent personality from which autonomous decisions can flow), rather than on the action in question.

Second, they "reject authenticity...because [they] believe it is too demanding as a condition of autonomous action" (1986, p.273). If we take authentic actions to involve, as I have suggested above, cohesion with or reflection of the stable and enduring values that constitute the self, Faden and Beauchamp maintain that an authenticity condition would "render nonautonomous many choices that are worthy of respect as autonomous" (1986, p.266). They maintain that people often make perfectly legitimate, deliberate choices that depart from their stable values. People should be permitted to act out of character without their decisions being deemed unworthy of respect as autonomous. They also suggest that people may not have a coherent set of values underlying legitimate choices, and that reference to stable values might simply be irrelevant in certain

[5] As will be further elucidated below, Faden and Beauchamp want their account of autonomy to be focused on actions, not persons. It is for this reason that they speak of an *action* being autonomous, rather than, as is more common in the theoretical literature, a person being autonomous with respect to her action.

situations. We make many everyday choices, such as choosing a washing powder or a sandwich for lunch, which may have nothing do to with our stable values, and should certainly not be deemed problematic because of this.

3. Competence assessments and the role of an authenticity condition.

Faden and Beauchamp wish to construct a theory in which low-stakes, everyday decisions are not unduly restricted. This is a laudable and important aim. They are right to hold that holding such decisions to an authenticity requirement is inappropriate and excessively restrictive. However, this does not mean that we should abandon the condition of authenticity entirely. The condition of authenticity can provide an invaluable means of evaluating some decisions, namely high-stakes decisions where the patient might be faced with intersubjective disagreement from assessors. We can draw this out by looking at how competence assessments tend to function in medicine – and when competence is questioned in medical settings. We can then see how an appeal to authentic values might be required to supplement accounts of competence and autonomy in a medical context.

The concept of competence plays the same role in medicine that Faden and Beauchamp envision for their model of autonomy – it is designed to discern which decisions should be respected or honored in a medical setting.[6] There is also overlap in requirements – Faden and Beauchamp's condition of understanding, for example, corresponds to the conditions of understanding and appreciation in Thomas Grisso and Paul Appelbaum's benchmark MacArthur Competence Assessment Tool for Treatment (MacCAT-T).[7] I will thus treat these concepts, in terms of their role and purpose, as equivalent. When we look at how competence functions in medical settings, however, we can see that competence assessments are not usually initiated when the patient wishes to make a low-stakes decision. Rather, questions of competence generally only arise in cases where the patient desires to make a choice that is at odds with the decision that healthcare providers believe she should make (Varelius 2011), or, more specifically, when the patient wishes to choose a course of action which "in the opinion of the physician in charge, threatens his or her well-being" (Drane 1985, p.17).

Erich Loewy suggests that the validity of the patient's choices is questioned when we disagree with the choices or do not believe them to be in the patient's best interest because we "consider our

[6] More on this distinction below.
[7] Again, we will look at this in more detail below.

choices to be eminently reasonable, sane, and rational, and have trouble conceding such attributes to a choice in conflict with our own" (1988, p.55). Bernard Gert, Charles Culver and K. Danner Clouser furthermore contend that it is common practice to implicitly incorporate a condition of rationality into judgments of autonomy and competence in medical contexts by interpreting "understanding and appreciating in such a way that the patient who made a seriously irrational decision was understood to be showing that he either did not understand or did not appreciate the consequences of his decision" (2006, p.227), and that this can result in a patient being found competent when he consents to a physician's proposed course of action, but incompetent when he refuses. It is thus important, when we are formulating a model of competence or autonomy for use in medical ethics, that we ensure decisions, particularly decisions in which a proposed course of action does not seem to correspond with the assessor's view of what is in the best interests of the patient, are not simply overridden in the face of intersubjective disagreement. We should pay careful attention to conditions of understanding and appreciation to ensure that intersubjective[8] notions of rationality or wellbeing are not smuggled into these standards.

It is here that an authenticity condition can be of use. The individually focused, value-neutral accounts of autonomy in theoretical philosophy can be adapted to this context, to ensure that autonomy or competence can be evaluated without relying upon intersubjective judgments. Rather than operating, as Faden and Beauchamp suggest, as an extra condition that decisions must meet, authenticity should be conceived as providing an underlying frame of reference that allows us to assess whether a decision is, for example, adequately understood or appreciated. By determining whether a decision corresponds to the enduring values of the agent, we can determine whether there are problems with these conditions without bringing the judgments of others into the equation. An underlying notion of authenticity can bolster and insulate the conditions of autonomy or competence in medical contexts against judgments from others that a decision, based on its content, must be irrational, or against the patient's best interests, and thus incompetent or nonautonomous. I will draw out the need for, and precise role of, such an underlying framework in the subsequent two sections.

There are two additional reasons that reference to authenticity in medical models of autonomy protects against intersubjective judgments, particularly judgments concerning wellbeing. The

[8] I use the term "intersubjective" rather than "external" or "objective" here and in what follows because I want to stress that there is no *objective* standard by which the *content* of someone's values and goals can be criticized – when we criticize the content of others' values and goals, we are doing so from our own conception of rationality, wellbeing, etc., or from standards that have achieved widespread intersubjective agreement in our society (see Wolf 1987).

notion of the Lockean self, consisting of enduring values and desires, which forms the foundation of the authenticity condition, provides support for the patient's ability to choose what is best for his future self, challenging the validity of certain types of intersubjective judgments concerning wellbeing. Such a theory emphasizes the connection between the present and future self, suggesting that a patient may be in the best position to judge what is best for his future self – based on his enduring desires and values. This gives us grounds for establishing a clear delineation concerning the role of the doctor; based on his superior technical knowledge, the doctor might be able to challenge a patient's belief that a certain course of action will lead her to achieve her goals. But the patient should be regarded as in a better position to make judgments about the goals themselves.

Furthermore, the theoretical background of authenticity-based autonomy provides us with a clear reason that respecting these decisions is so important. As noted above, in respecting the decisions that form a reflection of the values that constitute the person, we are given a concrete way of expressing the abstract value of respect for persons. To fail to respect these decisions is to fail to show adequate respect for the person. The clear link between respect for persons and authentic decisions provides another layer of protection by giving a clear account of their value, and of what respecting them means.

Though it is generally accepted that requiring that high-stakes decisions meet a condition of authenticity will involve greater restriction of decisions (Drane 1984; Faden and Beauchamp 1986), an authenticity condition, in providing a framework for evaluation which is independent of intersubjective judgments, in challenging the validity of certain types of intersubjective judgments concerning wellbeing, and in providing an imperative to respect certain self-regarding actions regardless of our feelings about them, has the potential to protect more high-stakes decisions from being overridden on the grounds that they do not meet standards of autonomy or competence. Faden and Beauchamp's point about authenticity and everyday decisions still stands, however. As authenticity should not be a requirement of all decisions, this should not be appealed to in the case of low-stakes decisions – this can be achieved simply by incorporating authenticity requirements into standards of competence that are only used as an evaluative tool when the decisions of the patient are called into question – as we have seen above, this tends to only be in cases where the assessors believe that the decision will threaten the patient's wellbeing. Alternatively, one could endorse a "sliding scale" model of competence along the lines proposed by Jim Drane (1984) and Allen Buchanan and Dan Brock (1986), where competence requirements are linked to the potential

risk involved in the decision. In this framework, authenticity-based requirements would only form part of an assessment of competence or autonomy for high-risk procedures.

4. Bolstering the condition of understanding

To draw out the need for an underlying notion of authenticity when assessing high-risk decisions, I will turn to Faden and Beauchamp's condition of understanding, or more specifically, their treatment of false beliefs.[9] I will show that without an underlying notion of authenticity, Faden and Beauchamp's condition of understanding is not adequately insulated against intersubjective judgments when it comes to high-stakes decisions that go against the assessor's notion of what is in the patient's best interests. Faden and Beauchamp provide a sophisticated and detailed account of the conditions under which a person understands the nature and implications of their actions. They suggest that, in the sense important to us here, the question of whether someone understands what they are doing can generally be understood as asking whether they have *justified beliefs* about the *consequences* of what they are doing.

Their specification that patients must have justified beliefs *about the consequences* of what they are doing has potential to correspond to the delineation I have outlined above. Perhaps patients are only required to be aware of the foreseeable consequences of their chosen course of action, in order to ensure that their proposed decision is the one that is likely to lead them to achieve their goals, while the goals themselves are not scrutinized as part of the requirement of understanding. However, when we turn to Faden and Beauchamp's treatment of false beliefs, we can see that this distinction is not adequately established, revealing the need to bolster this account through reference to authentic values.

Faden and Beauchamp contend that "[t]o the extent that a person's understanding is based on false beliefs about that which would otherwise be relevant to an understanding of the action, performance of that action is less than fully autonomous" (1986, p.253). They note, however, that the challenge of distinguishing true from false beliefs, and the inherent uncertainties attending decision-making, present major problems when assessing autonomy. They suggest that, rather than requiring that a decision is based on a "true" belief, we should use an evidential standard for assessing the acceptability of beliefs. This will involve asking whether a person, on the basis of the

[9] This approach might also show promise for Faden and Beauchamp's condition of noncontrol - they themselves suggest that a condition of authenticity shows the most promise when it comes to adequately capturing both external and internal controlling influences (1986). However, an adequate assessment of this complicated concept is beyond the scope of this paper.

available evidence, is *justified* in holding the belief that motivates her action. They thus shift their standard from true belief to justified belief. They claim, however, that when "the justifiability of a belief or the warrant for its assertion is inherently and unavoidably contestable, there may be no adequate grounds for determining whether a given belief compromises understanding" (1986, p.254). They suggest that although this invokes legitimate epistemological problems that are relevant to the discussion at hand, this should not prevent us from adopting their proposed justified belief standard for assessing understanding.[10]

Faden and Beauchamp's evidential standard for justified belief is undoubtedly superior to a true belief standard when it comes to dealing with the inherent uncertainties and probabilities involved in calculating the outcome of a certain course of action. We need a standard of autonomy that allows for unexpected outcomes, even when people are acting upon the best available evidence. Their proposed standard, however, does not give us a means of making important distinctions between what they refer to as "contestable" beliefs. To reveal this, we can turn to two examples that Faden and Beauchamp give of contestable beliefs:

> 1) "If I consent to this blood transfusion I will burn in hell."
> 2) "I know that 50% of patients with my problem do not survive this type of surgery. I am a fighter, and if I consent to this type of procedure, I will survive." (1986, p.254).[11]

They note that these beliefs could be central to decision-making in a medical context, but that they "may be viewed by others, including those seeking consent, as highly questionable, poorly reasoned, or patently false" (1986, p.254). Faden and Beauchamp give us no further indication of whether these beliefs are different in any way, and whether either or both of them should be regarded as ultimately undermining autonomy. Leaving this open or noting that the decision may be impossible is insufficient given that in practical contexts, a decision will need to be made.

[10] The so-called "Gettier Problem" reveals just how vexing some of these issues can be; see Gettier (1963).

[11] Faden and Beauchamp also provide a third example here: "Nothing good can come of my consenting to this procedure, because no matter what new skills and coping styles I might develop, I will never want to live as a quadriplegic." This is a particularly difficult case, and one that has attracted considerable attention in existing literature (see e.g. Savulescu 1994; Cowart and Burt 1998). Without getting into the details of this nuanced issue, I believe such a case can be satisfactorily dealt with by my proposed approach, by first making an effort to dispel unrealistically negative conceptions of what it means to be a quadriplegic (see Savulescu 1994), but, after reasonable efforts have been made, accepting the patient's decision to refuse treatment on the basis of his enduring values.

This problem is potentially compounded when we take into account that this justified belief standard is designed to capture "the common-sense conception of reasonable…belief and assertion that underlies ordinary social agreements about what is veridical" (1986, p.254). There seems at least a risk here, in the absence of further indicators, that the justifiability of the beliefs in these cases will ultimately end up being determined by whether the assessor feels they are reasonable, or perhaps even based on whether the outcome of the decision corresponds to the assessor's desired outcome. Given the well-established propensity for autonomy and competence assessments to ultimately fall back on intersubjective judgments, we must be particularly careful to avoid this risk.

There is an important sense in which the beliefs in Faden and Beauchamp's two scenarios are distinct, and they can be distinguished if we make reference to authentic values when evaluating these beliefs. The contestable belief in the first scenario is based upon the enduring values of the patient; the (presumably religious, Jehovah's Witness-based) belief system that he has shaped his life around. This is the type of contestable belief that should therefore be out of range of the assessor's scrutiny. The belief in the second scenario, however, is about the probability of a certain outcome obtaining. Holding a false belief about the probabilities of the outcomes, in this scenario, could prevent the patient from achieving his goal in consenting to the procedure (we can infer, from the information given above, that his goal here is survival). We therefore have an important distinction between these two cases – in the first case, the patient has correctly identified the means to achieving his own goals based on his own value system; in the second case, the patient holds a belief that may interfere with achieving his own goals based on his own value system. Reference to the enduring beliefs, desires and goals of the patient allows us to identify what is wrong with the second false belief, and prevents interference with the first patient's conduct on the basis of his contestable belief. Faden and Beauchamp suggest elsewhere that a refusal of a blood transfusion from a Jehovah's Witness should not be overruled on the grounds that this action does not qualify as autonomous (and express concern that an authenticity condition, albeit conceived more stringently than what I have proposed, might lead to this result). However, their model, in simply identifying the underlying belief of the Jehovah's Witness as contestable, with possibly no adequate grounds for determining whether it is justified, leaves this action vulnerable to being overridden on the grounds that the underlying belief is not shared by the assessor.

An authenticity-based conception of understanding allows us to place the enduring values and beliefs that people shape their lives around out of reach of scrutiny, and to identify problematic

beliefs; those that obstruct people from pursuing their subjectively-chosen goals. Rather than scrutinizing the enduring beliefs and values that constitute the self on the grounds of justifiability, we should view these values and beliefs as a *basis* for the justification of beliefs concerning treatment. To make these self-forming beliefs, desires and values candidates for scrutiny on the basis of justifiability is to abandon an individual framework upon which a meaningful notion of autonomy can be based, and necessarily to adopt a standard of justification which must rest on something outside one's own individual values. If individual values themselves are appropriate targets for scrutiny, it is difficult to see how we can avoid the risk of intersubjective judgments forming the basis for this evaluation. It also puts doctors in the position of deciding about the acceptability of patients' subjectively held values; a task for which they have no special claim to authority or expertise.

5. Authenticity in Grisso and Appelbaum's concept of competence

We can see further advantages of an appeal to authentic values as a means of assessing competence or autonomy by turning to Grisso and Appelbaum's concept of competence. Their formulation of competence provides the basis for their aforementioned MacCAT-T, which is widely regarded as a "gold standard" (Vollmann 2006, p.289; see also Swindell 2009) for competence assessments. Grisso and Appelbaum are concerned about the tendency to invoke a notion of rationality to classify decisions that are seen by others as "unconventional or not in the patient's best interest" (1998, p.53) as incompetent. They contend that the unpopularity or eccentricity of a choice should not be, on its own, enough to disqualify it from being respected; to do otherwise, they suggest, would be inconsistent with the value that society places on autonomy, and one's ability to make one's own choices. Their model of competence is thus carefully structured to protect the content of choices themselves against intersubjective judgments.

Grisso and Appelbaum propose four conditions for competence; ability to communicate a choice, understanding, appreciation, and reasoning. While Faden and Beauchamp's condition of understanding is designed to include issues of appreciation, Grisso and Appelbaum have separate conditions for understanding, which involves understanding the meaning of the information disclosed, and appreciation, which involves seeing the relevance of this information to one's own circumstances. The patient in Faden and Beauchamp's second scenario, for example, meets Grisso and Appelbaum's condition of understanding, because he understands the information that has been conveyed to him, but he exhibits a failure of appreciation, because he has not applied this information to his own situation, or appreciated the significance of this information in relation to

his own circumstances. Because the issues pertaining to false beliefs are examined under Grisso and Appelbaum's discussion of appreciation, we will focus on this criterion in what follows.

Grisso and Appelbaum's concept of appreciation is primarily concerned with false beliefs. A failure of appreciation, under this notion, results from holding patently false beliefs – which they take to mean substantially irrational, unrealistic, or amounting to a considerable distortion of reality. They stress that this criterion is designed to assess the beliefs or premises upon which choices are based, and not the choices themselves. They use this approach to distinguish between a woman who refuses amputation because she does not believe her foot is gangrenous (which can be seen as a belief that obstructs her ability to achieve her goals, akin to the situation in Faden and Beauchamp's second scenario), and a man who refuses life-extending cancer treatment to spend his remaining time with his grandchildren, who he values above all else. An appeal to underlying values here shifts the focus from the choice itself, which may seem odd, wrong or dangerous to an outsider, to the reasons for the choice; namely, seeing the choice as a means to achieve one's deeply held values. The relation between the choice and the values that underlie it reveal that the choice itself, which may seem to indicate that something is wrong with the process leading up to it, is in fact not made on the basis of false beliefs – the choice represents an appropriate means of achieving the patient's specified ends.

Grisso and Appelbaum also expressly address cases in which patients desire a course of action, against physician recommendations, based on religious values. Here, they suggest that an appraisal of competence should be limited only to ascertaining that the religious beliefs are genuinely held. They suggest that this can be assessed by ascertaining whether these beliefs are enduring (i.e. whether they predate the treatment decision) and "whether the patient has previously behaved in ways consistent with these beliefs" (1998, p.48). This gives us an indication not just of how authenticity might be used to assess and protect religiously-motivated actions, but how we can ascertain whether beliefs are authentically held in practical situations. It also gives us a method of evaluating authenticity which is not, as Faden and Beauchamp feared, unduly concerned with the global structure of desires that form the hallmark of the autonomous person. By referring only to the enduring values that motivate the desired course of action, and by focusing only on establishing that these values are indeed enduring, Grisso and Appelbaum's system of evaluation places the focus squarely on the choice at hand.

It is important to note here, that although Grisso and Appelbaum's model is strengthened insofar as it appeals to authentic values, they should not necessarily be taken to be advocating a hard distinction between one's deeply held values (out of reach of scrutiny) and beliefs about how one can best achieve these values (potentially problematic and autonomy compromising). They suggest that if the loving grandfather was basing his decision on other factors, we might have more of a problem regarding it as competent. They also do not base the religious believer's actions purely on an assessment of whether his values are enduring and consistent with other choices; they suggest that whether this is a legitimate religious sect (i.e. whether other people hold these values) might play a role in determining legitimacy too. Perhaps the inclusion of this factor only comes into the equation due to the necessity of finding factors that can be assessed – exactly what position they take on the extent to which deeply held values might be legitimately scrutinized is unclear. I think that explicitly endorsing the clear distinction I advocate, for the reasons I have outlined above, would strengthen their account and better allow them to achieve their stated goal of protecting choices from being overridden based on intersubjective disagreement concerning their content. This seems to be the approach that Allen Buchanan and Dan Brock take in their concept of competence, where, upon considering a Christian Scientist's refusal of surgical intervention, they argue that while "these values are highly unusual, if they have had a central and enduring place in his life plan, they would make sense of his choice and indicate no decisionmaking deficit" (Brock 1991, p.10). But regardless of the extent to which my contentions concerning this matter might disagree with some aspects of Grisso and Appelbaum's model, their appeal to authentic values as a basis upon which choices can be assessed reveals an advantage of an underlying notion of authenticity – it gives us an independent process by which decisions can be evaluated, shifting focus from the content of choices, and making it less likely that choices will be overridden based on intersubjective rejection of the desired outcome. They also give us an indication of how the authenticity of values themselves can be ascertained in clinical contexts.

6. Is this enough?

I have argued that reference to authentic values is valuable in medicine because it better protects decisions from being overridden on intersubjective grounds. Relying on authentic values as the ultimate justificatory basis for the decisions that we should respect and honor in a medical context, however, potentially runs into problems when we are faced with certain desires. One might suggest that an authenticity requirement, as I have construed it, is not stringent enough when it comes to high-stakes decisions that compromise intersubjectively conceived notions of wellbeing – that some problematic decisions will meet the requirement of correspondence with stable and enduring

values. An anorexic or depressed patient's decision to refuse treatment, for example, seems, on the face of things, capable of meeting the condition of authenticity; it may be based on stable, enduring, and even coherent values. Patients with these conditions are also often capable of making highly rational decisions, and may thus be unlikely to exhibit deficits in Appelbaum and Grisso's criteria of competence. At the same time, there seems to be something problematic about endorsing these decisions. I will consider each of these cases in turn.

Anorexic patients, with their single-minded focus on achieving their goals, and their enduring values and beliefs corresponding to a desire to starve themselves, sometimes experienced as part of their identity (Tan et al 2006), can be regarded as highly autonomous, and as thus potentially posing a severe problem for authenticity-based accounts of autonomy. Jacinta Tan and colleagues challenge the idea that Grisso and Appelbaum's concept of competence can adequately deal with some anorexic patients' refusal of treatment; they contend that some anorexic patients can meet the requirements of competence, and that we should thus revise the notion of competence to exclude "pathological values". Because the distinctive values of the anorexic person are new to the person since the development of the disorder, are closely connected with the disorder, and can in some cases be "unusual or bizarre" (Tan et al 2006), presumably from an intersubjective standpoint, Tan and colleagues contend that they cannot be counted as authentic to the patient.

It is worth considering the extent to which the approach I have thus far suggested could be thought to deal with the problems raised by Tan and colleagues. The fact, for example, that new values have arisen since the advent of the disorder could be thought to interfere with the enduring values of the anorexic patient in a way that might compromise autonomy. Some patients feel alienated from these desires, or they generate conflict with other enduring desires that form the self (Tan et al 2006). However, other patients come to identify with these desires and incorporate them into a coherent personality structure. When these desires have endured for long enough, they could be said to form a part of the enduring values and desires that constitute the anorexic patient's Lockean self. If these desires are indeed incorporated into the self in this way, any decision based upon them, including the decision to refuse treatment, will count as authentic; as flowing, in the appropriate way, from the self. An appeal to enduring values will disqualify some subjects, however, the desires of others will be regarded as autonomous in the sense I have advocated, and thus unable to be overridden on these grounds.

Grisso and Appelbaum also point out, in a response to this study, that many of the subjects involved in fact seem to show deficits in reasoning and appreciation (2006). To focus, as above, on appreciation, quotes from the subjects of the study do seem to indicate a pervasive problem here. Several participants noted that although they understood the risks of death and disability related to anorexia, but stated "[p]art of me didn't believe it" (Tan et al 2006, p.72); "I didn't think it applied to me at all" (p.71) and "I won't die" (p.74). This indicates that failure of appreciation presents a problem for some anorexic patients, leading to problems with competence. However, it is again plausible that this might not affect all anorexic patients – some desires based on anorexic values are likely to remain autonomous. Though these considerations do a better job of pointing out problems with some anorexic desires than Tan and colleagues contend, it is clear that they will not achieve their goal of rendering all such desires nonautonomous.

Much the same can be said about the refusal of treatment from someone with depression. Like many desires stemming from anorexia, depressive desires also seem, in some cases, to be distinguished by their enduring and pervasive nature. In addition, and in contrast to many other mental illnesses, the patient's capacity to offer rational explanations for her choices is often not compromised by depression (Sullivan and Youngner 1994). In this case, as above, one may argue that decisions that would qualify as competent on the basis of the framework I have advocated should not be considered authentic, because they stem from the "pathological values" of depression (see Charland 2006).

There may be a possibility of highlighting a problematic element in these sorts of decisions without criticizing the values that motivate them directly. In his analysis of depression and competence to refuse treatment, Abraham Rudnik, reluctant to place the values that motivate decisions directly in range of scrutiny, suggests that we might deal with depressive desires through a strengthened condition of coherence. Concerned, like Tan and colleagues, with the significant changes in preferences and values that depression can precipitate, he suggests that we should "compare pertinent preferences of the individual during depression to preferences regarding the same subject matter held by that individual when not depressed" (2002, p.153). Although this approach is attractive in some respects, we should be wary of requiring a strong degree of coherence in the desires that constitute the self. As Rudnik himself notes, most individuals are not likely to meet a strong requirement that their preferences, values and desires cohere, this might rule out change of preferences, and there is no good measure of strong coherence (2002). This is also likely to shift the focus from the decision at hand back to the global structure of the patient's desires; something

that Faden and Beauchamp sensibly caution against. In any case, this is not likely to render all depressive desires to refuse treatment nonautonomous; where depression is ingrained and enduring, the patient could likely meet this condition of coherence, even judged over time (Rudnik 2002).

As with anorexia, it is plausible that some depressive desires could be deemed nonautonomous due to a failure to meet Grisso and Appelbaum's four conditions of competence. Putting depression with psychotic features aside, there is widespread agreement that some subjects with depression will exhibit deficits in appreciation, either due to a failure to appreciate the relevance of information concerning the mental illness to one's own circumstances, or due to an inability to comprehend the personal consequences of one's choice (Hindmarch et al 2013). This is supported by empirical evidence; a study conducted by Grisso and Appelbaum found that, where a depressed patient exhibited a deficit in competence, appreciation was the most common problem (1995). However, they also found that the majority of patients with depression exhibited no deficit at all, a finding that has since been backed up by other studies (Hindmarch et al 2013).

A notion of competence resting on the basis of authenticity, then, will not judge all anorexic and depressive desires to refuse treatment as nonautonomous. However, we should not be so quick to fall back on a notion of pathological values. The idea that values should be ruled incompetent in virtue of the fact that they stem from a mental disorder creates more problems than it solves, particularly when incorporated into a notion of competence that is designed to apply to a wide spectrum of clinical situations. In a study of 2100 physicians of various ages and specialties, Markson and colleagues reported that the majority incorrectly believed that certain mental disorders (such as dementia and psychosis) establish incompetence (1994). An adequate notion of competence should be able to recognize that patients with mental illness can and do make competent decisions that should be respected. Introduction of pathological values risks reinforcing the already prevalent tendency to equate mental illness with incompetence. Appelbaum and Grisso also express reservations regarding altering the model of competence in response to one specific set of mental disorders – warning that this could have negative or unwanted impacts when the standards are applied to people with other categories of mental disorders (2006). It is also important to note, particularly in the case of depression, that symptoms and thus impairment fall on a spectrum. Although it is plausible to hold that in some severe cases, impairment may exist, it is not plausible to hold that impairment exists in every case, even when it comes to high-stakes refusals, such as the refusal of lifesaving treatment. Mark Sullivan and Stuart Youngner suggest

that these issues of degree are best captured by maintaining "a clear distinction between the diagnosis of depression and assessment of competence" (1994, p.977).

There is a further possibility here, that does not involve directly labelling certain values pathological and thus autonomy-compromising on an ad hoc basis. We could introduce a 'weak substantive' condition, along the lines suggested by Paul Benson (2005). Benson maintains that, when faced with problematic desires that meet the conditions of content-free autonomy, the optimal answer is not to place direct restrictions on which desires can count as autonomous. Instead, he suggests we can find a middle ground; an incorporation of some normative content into a theory of autonomy without resorting to restriction of certain desires, namely, the introduction of a requirement that agents have a sense of their own competence and worth as decision-making agents. This is attractive not just because it avoids direct restrictions on content. It can also be justified without contravening the value of individualistic, authenticity-based autonomy. I noted above that respecting authentic desires is so important because doing so expresses the value of respect for persons. When an agent wishes to act in a way that does not reflect a sense of her own worth, she is not showing sufficient regard or respect for herself. To facilitate this decision in a medical context could be seen as endorsing the agent's conception of herself as lacking worth, and on these grounds be at odds with the value of respect for persons. Refusing to facilitate such desires, at least, must not be seen as an offense against the value of respect for persons in the same way that a refusal to respect other authentic decisions is.

A weak substantive condition of autonomy seems to have potential when applied to the cases discussed above. Tan and colleagues noted that a common characteristic of their anorexic research subjects was a tendency to "devalue themselves and also life in general" (2006, p.73). This led to the research subjects expressing that they didn't care what happened to them, and that it wouldn't matter if they died. An inadequate sense of self-worth, leading to submissive and self-sacrificing behavior, is often recognized as a hallmark of anorexic patients (see Eth et al 1981). Depressive patients exhibit the same tendencies; low self-esteem and feelings of worthlessness are characteristic of depression (Nestler et al 2002, Rudnik 2002) which can lead patients to believe that suffering or death is deserved and should not be prevented (Sullivan and Youngner 1994). Benson's weak substantive condition for autonomy provides us with a way of dealing with these tendencies without placing direct restrictions on the content of desires, and can allow us to avoid facilitating these decisions, which could be seen as endorsing this inadequate sense of worth.

However, even if we do accept this additional, weak substantive condition as a requirement of autonomy or competence in medical settings, we cannot ensure that no anorexic or depressive desires to refuse treatment, even life-saving treatment, will qualify as autonomous. This can only be achieved by direct restrictions on the content of the desires that qualify as autonomous or competent. Autonomy, conceived as based upon the authentic values of the patient, will never provide the perfect means of sorting all the desires and decisions that we wish to respect from the desires that we find problematic. At this point, however, rather than accepting direct, intersubjective restrictions on the content of desires or values, I contend that we are better off abandoning the idea that autonomy and competence should play this role in medical ethics. If we expect our notions of autonomy to fulfill this function, we invite placing paternalistic and intersubjective standards around the concept of autonomy, thus rendering it meaningless. For autonomy to be a meaningful notion in medicine, patients must be free to choose in ways that are not endorsed or approved by others.

The price of a meaningful concept of autonomy is to accept that people can autonomously or competently make some decisions that we might not like. We then have two options. We can either honor these decisions, or we can question the idea that autonomy or competence must trump all other values when assessing whether to honor decisions in medical ethics. I contend that the second option is a preferable means of dealing with anorexic or depressed desires, compared to attempting to directly exclude them from our notion of autonomy or competence. To entertain the notion that competent decisions might sometimes be justifiably overridden is very unpopular in the literature on competence (Brock 1991; Skene 1991; Culver and Clouser 2006). The reasons for this are clear; to accept that competent decisions may be overridden in some circumstances contravenes established legal standards, and opens up the possibility for abuse. But the desire to uphold this requirement has led to the development of models of autonomy which are upheld as of paramount importance, but that are structured around intersubjective acceptability. That is, when we accept that autonomous decisions can never be overridden, and we do not wish to honor a certain decision, we can only do so by deeming this decision nonautonomous. When we use our disapproval of a certain course of action to delineate what counts as an autonomous decision, we are simply paying lip service to the value of respect for autonomy. We would likely see a greater ability to exercise autonomy in a meaningful sense if we endorse an authenticity-based notion of autonomy which might sometimes (according to strict guidelines) be justifiably overridden due to overtly paternalistic considerations, than we are when intersubjective judgments concerning wellbeing are incorporated into standards of autonomy or competence, with every decision that

does not correspond to these requirements ruled nonautonomous or incompetent (see White 2017).[12]

Conclusion

The condition of authenticity, the centerpiece of many theoretical philosophical models of autonomy, also has a role to play in concepts of competence and autonomy in medical ethics. I have advocated that authenticity be seen as playing a certain role in medical autonomy or competence assessments; rather than an additional criterion which decisions must meet, it should be seen as an underlying framework which better allows us to assess the existing criteria of autonomy and competence. This gives us a means of distinguishing between contestable beliefs that patients shape their life around and that determine a patient's goals, and beliefs that may present an obstruction to achieving these goals. This approach protects the individual values and beliefs that form the basis of the self, while better identifying the problematic false beliefs that may undermine autonomy. I have thus attempted to display, against the prevailing view, that appropriate reference to authentic values in autonomy and competence assessment can lead us to honor more of the patient's decisions, particularly high-stakes decisions that are likely to be challenged on intersubjective grounds due to concern for wellbeing. I have considered certain problematic decisions that can meet the criteria of authenticity-based autonomy or competence as I have conceived it, and suggested some possibilities for dealing with these sorts of desires, ultimately concluding that a meaningful notion of autonomy is more important than a notion that corresponds to our beliefs about which decisions should be honored. A final note: in drawing out these arguments, I have relied on a highly general account of authenticity; whether simply being enduring is enough for a desire to qualify as authentic, or whether and to what degree we might sensibly incorporate some sort of requirement of coherence or satisfaction, while retaining a sufficiently permissive and choice-focused standard, is unfortunately beyond the scope of this paper. I hope, however, that in outlining the role for authenticity in medical ethics I have displayed not just the need to incorporate this consideration into practical models of autonomy, but also that developing accounts of autonomy designed for these purposes might be a fruitful endeavor for autonomy theorists.[13]

[12] This need not fundamentally change the purpose of autonomy or competence requirements; we could view these decisions as prima facie worthy of special respect and protection, but able to overridden in exceptional circumstances. Indeed, this is the role that Faden and Beauchamp envision for autonomous decisions.

[13] Many thanks to Dietmar Hübner and two anonymous referees from the HEC Forum for their useful comments on an earlier draft of this paper.

References

Benson, P. (2005). Feminist intuitions and the normative substance of autonomy. In J. Taylor (Ed.), *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy* (pp. 124-142). Cambridge: Cambridge University Press.

Bratman, M. (2007). Planning agency, autonomous agency. In Bratman, M. *Structures of agency* (pp.195-221). New York: Oxford University Press.

Brock, D. (1991) Decisionmaking competence and risk. *Bioethics*, 5(2), 105-112.

Buchanan, A. and Brock, D. (1986) Deciding for others. *The Milbank Quarterly*, 64(Suppl.2), 17-94.

Charland, L. (2006). Anorexia and the MacCAT-T test for mental competence: validity, value and emotion. *Philosophy, Psychiatry and Psychology*, 13(4), 283-287.

Cowart, D. and Burt, R. (1998). Confronting death: who chooses, who controls? *The Hastings Center Report*, 28(1), 14-24.

Drane, J. (1984). Competency to give an informed consent. *Journal of the American Medical Association*, 252(7), 925-927.

Drane, J. (1985). The many faces of competency. *Hastings Center Report* (15)2, 17-21.

Dworkin, G. (1988) *The theory and practice of autonomy*. Cambridge: Cambridge University Press.

Ekstrom, L. (2005). Autonomy and personal integration. In J. Taylor (Ed.), *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy* (pp. 143-161). Cambridge: Cambridge University Press.

Eth, S., Eth, C., and Edgar, H. (1981). Can a research subject be too eager to consent? *Hastings Center Report*, 11(1), 20-21.

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

Frankfurt, H. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5-20.

Gert, B., Culver, C., and Clouser, K. (2006). *Bioethics: A systematic approach*. Oxford: Oxford University Press.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 23(6), 121-123.

Grisso, T. and Appelbaum, P. (2006). Appreciating anorexia: decisional capacity and the role of values. *Philosophy, Psychiatry and Psychology*, 13(4), 293-297.

Grisso, T. and Appelbaum, P. (1998). *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. Oxford University Press, New York.

Grisso, T. and Appelbaum, P. (1995). Comparison of standards for assessing patients' capacities to make treatment decisions. *American Journal of Psychiatry*, 152(7), 1033-1037.

Hindmarch, T., Hotopf, M. and Owen, G. (2013). Depression and decision-making capacity for treatment or research: a systematic review. *BMC Medical Ethics*, 14(54), 1-10.

Loewy, E. (1988). Changing one's mind: when is Odysseus to be believed? *Journal of Geriatric Internal Medicine*, 3(1), 54-58.

Markson, L., Kern, D., Annas, G. and Glantz, L. (1994). Physician assessment of patient competence. *Journal of the American Geriatrics Society*, 42(10), 1074-1080.

Nestler, E., Barrot, M., DiLeone, R., Eisch, A., Gold, S. and Monteggia, L. (2002). Neurobiology of depression. *Neuron*, 34(1), 13-25.

Noggle, R. (2005). Autonomy and the paradox of self-creation. In J. Taylor (Ed.), *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy* (pp. 143-161). Cambridge: Cambridge University Press.

Rudnik, A. (2002). Depression and competence to refuse psychiatric treatment. *Journal of Medical Ethics*, 28(3), 151-155.

Savulescu, J. (1994). Rational desires and the limitation of life-sustaining treatment. *Bioethics*, 8(3), 191-222.

Skene, L. (1991). Risk-related standard inevitable in assessing competence. *Bioethics*, 5(2), 113-117.

Sullivan, M. and Youngner, S. (1994). Depression, competence, and the right to refuse lifesaving medical treatment. *The American Journal of Psychiatry*, 151(7), 971-978.

Swindell, J. (2009). Two types of autonomy. *The American Journal of Bioethics*, (9)1, 52-53.

Tan, J., Stewart, A., Fitzpatrick, R. and Hope, R. (2006). Competence to make treatment decisions in anorexia nervosa: thinking processes and values. *Philosophy, Psychiatry and Psychology*, 13(4), 267-282.

Taylor, J. (2005). Introduction. In J. Taylor (Ed.), *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy* (pp. 1-29). Cambridge: Cambridge University Press.

Varelius, J. (2011). Decision-making competence and respect for patient autonomy. *Res Cogitans*, 8(1), 33-42.

Vollmann, J. (2006). "But I don't feel it": values and emotions in the assessment of patients with anorexia nervosa. *Philosophy, Psychiatry, and Psychology*, 13(4), 289-291.

White, L. (2017). How autonomy can legitimate beneficial coercion. In J. Gather, T. Henking, A. Nossek and J. Vollmann (Eds.) *Beneficial coercion in psychiatry? Foundations and challenges* (pp.85-99). Münster: Mentis.

Wolf, S. 1987. Sanity and the metaphysics of responsibility. In F. Schoeman (Ed.) *Responsibility, character and emotions: new essays on moral psychology* (pp. 46-62). New York: Cambridge University Press.