Authenticity in Bioethics: Bridging the Gap between Theory and Practice

Jesper Ahlin Marceta

Doctoral thesis in philosophy
KTH Royal Institute of Technology
Stockholm, 2019

Copyright information, etc.

Ahlin Marceta, J. (2019). *Authenticity in Bioethics: Bridging the Gap between Theory and Practice* (doctoral thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

ISBN 978-91-7873-124-4 ISSN 1650-8831

Included articles

1. Ahlin, J. (2018). The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent. *Medicine, Health Care and Philosophy* 21(1), 43–50
2. Ahlin, J. (2018). What Justifies Judgments of Inauthenticity? *HealthCare Ethics Committee Forum* 30(4), 361–377
3. Ahlin Marceta, J. (2018). A Non-Ideal Authenticity-Based Conceptualization of Personal Autonomy. *Medicine, Health Care and Philosophy* DOI:doi.org/10.1007/s11019-018-9879-1
4. Ahlin Marceta, J. Nine Cases of Possible Inauthenticity in Biomedical Contexts and What They Require from Bioethicists (manuscript)

Contact: jahlinmarceta.com

Typeset by the author. Printed by US-AB, Stockholm, Sweden.

Abstract

The aim of this doctoral thesis is to bridge the gap between theoretical ideals of authenticity and practical authenticity-related problems in healthcare. In this context, authenticity means being "genuine," "real," "true to oneself," or similar, and is assumed to be closely connected to the autonomy of persons. The thesis includes an introduction and four articles related to authenticity. The first article collects various theories intended to explain the distinction between authenticity and inauthenticity in a taxonomy that enables oversight and analysis. It is argued that (in-)authenticity is difficult to observe in others. The second article offers a solution to this difficulty in one theory of authenticity. It is proposed that under certain circumstances, it is morally justified to judge that the desires underlying a person's decisions are inauthentic. The third article incorporates this proposition into an already established theory of personal autonomy. It is argued that the resulting conceptualization of autonomy is fruitful for action-guidance in authenticity-related problems in healthcare. The fourth article collects nine cases of possible authenticity-related problems in healthcare. The theory developed in the third article is applied to the problems, when this is allowed by the case-description, to provide guidance with regard to them. It is argued that there is not one universal authenticity-related problem but many different problems, and that there is thus likely not one universal solution to such problems but various particular solutions.

Keywords: Authenticity, autonomy, decision-making, healthcare, paternalism, informed consent, bioethics

Contents

Acknowledgments

Firstly, I want to express my warmest gratitude to my main supervisor Barbro Fröding and my assistant supervisors Niklas Juth and Sven Ove Hansson. Throughout these years, Barbro has encouraged me to explore my philosophical interests while helping me keep my eyes on the ball. Her supervision has been invaluable. I am deeply grateful for her commitment to my research, education, and well-being. Whereas Barbro has enabled me to always see the bigger picture, Niklas's deep knowledge of the subject has forced me to think carefully. It was his research and detailed comments that first made me aware of the problem that I now claim to have solved in this thesis. I am also indebted to Sven Ove for his accurate remarks on my writing, his willingness to teach and discuss philosophy, and for giving me the opportunity to join *Theoria* as the journal's editorial assistant under his leadership.

At KTH, I have had the privilege to work with Patrik Baard, William Bülow, Robert Frisk, Henok Girma, Karim Jebari, Jesper Jerkert, Björn Lundgren, Payam Moula, Maria Nordström, Karl Sörensson, Anna Wedin, and Li Zhang in different stages of their PhD educations. My gratitude is extended to them, and to the faculty at the Division of Philosophy, namely John Cantwell, Karin Edvardsson Björnberg, Till Grüne-Yanoff, Mikael Karlsson, Niklas Möller, Tor Sandqvist, Linda Schenk, and Per Wikman Svahn. Tor's dedication in leading the higher seminar has been exemplary, and as head of department John is truly worthy of recognition for his devotion to, and trust in, the PhD students. Furthermore, I am happy to have shared my time here with Johan Berg, Helena Björnesjö, Hana Möller Kalpak, Niklas Norberg, Martin Rissler, Marcus Widengren, Edvin Åström, and Sebastian Östlund from the TaMoS staff. And, of course, my education and teaching have built on the administrative groundwork of Betty Jurdell and Fatemeh Tayebi.

Many of the ideas in this thesis have been discussed at the CHE seminar at Karolinska Institutet and at the biannual FORTE workshops in Stockholm, Gothenburg, and Linköping 2015–18. Here, I would particularly like to thank Leila Alti, Sara Belfrage, Helene Bodegård, Greg Bognar, Gert Helgesson, Petter Karlsson, Niels Lynöe, Christian

Munthe, Tomas Månsson, Lars Sandman, and Manne Sjöstrand for their input. Further to that, I am especially grateful to Gert for reviewing first my licentiate thesis and later the present doctoral thesis.

I also wish to thank Ulrik Kihlbom for his critical but constructive remarks on my previous work, and my many inspiring teachers; Marcus Agnafors, Martin Andersson, Martin Berzell, Erik Carlson, Lennart Göth, Helen Lindberg, and Lars Lindblom, to mention but a few names.

Finally, I am thankful to my family for their unconditional support throughout my education; my mother Helene, my father Örjan and his wife Maria, my siblings Ingrid, Martin, and Patrik and his Ida, and my grandparents Rolf, Barbro, Rune, and Märit. I also wish to thank my parents-in-law, Jovica and Natasa, and Kalle, Robert, and Ivan. But most importantly I am grateful to my wife Stella, whose endless patience and encouragement has made this work possible.

This research was supported by the Swedish Research Council (Vetenskapsrådet) and the Swedish Research Council for Health, Working Life and Welfare (FORTE), contract NO. 2014–4024, for the project Addressing Ethical Obstacles to Person Centred Care. I am also thankful to SciencesPo and the CEVIPOF institute in Paris, France, for hosting me as a visiting PhD student for one month in 2018, and to the Swedish Institute for granting me an apartment during my stay.

Stockholm, January 2019 *Jesper Ahlin Marceta*

Part I

Introduction

Authenticity in Bioethics

Respect for autonomy is a central moral principle in bioethics. The term autonomy comes from the ancient Greek *auto*, which means "self," and *nomos*, which means "law." Being autonomous means that one is self-governing. In biomedical contexts various concepts are associated with the concern of patients' autonomy, perhaps most notably *decision-making capacity* and *voluntariness*. That is, a patient is less autonomous to the extent that she lacks decision-making capacity and to the extent that she is not acting or choosing voluntarily. Sometimes, authenticity is also invoked; a patient is less autonomous to the extent that her actions or choices are inauthentic, or so the idea goes (cf., e.g., Christman 2009). The aim of this thesis is to make theoretical ideals of authenticity helpful in practical biomedical contexts, to further protect the autonomy of patients.

There are various uses of the term "authentic" in ordinary English. Lauren Bialystok identifies three main variations; in the first sense, authentic is synonymous with "original," as in "being continuous with a historical entity" (2014, p. 275). A 50's-style diner is authentic in this sense if it actually opened in the 1950s, was typical of that era, and has remained unchanged since then. In the second sense, it is synonymous with "real," as opposed to "fake." A citation is authentic in this sense if it reflects what the cited person actually said and has not been fabricated or distorted. In the third sense, authentic means "true to oneself" or "genuine." When a person is authentic in this sense, her behavior "converges with who she actually is" (p. 278). It is this sense of the term and how it relates to autonomy that is of interest here.

This thesis includes four articles about authenticity. The first article, entitled "The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent," collects theories that are intended to explain or conceptualize authenticity (Ahlin 2018a). In it, I argue that authenticity is difficult to observe in others. I call it "the Determining Authenticity article," or variations thereof. The second

article is entitled "What Justifies Judgments of Inauthenticity?" (Ahlin 2018b). In it, I formulate a proposal of how judgments that someone else's desires are inauthentic may be justified. I call it "the Inauthenticity Judgments article," or similar. In the third article, which is entitled "A Non-Ideal Authenticity-Based Conceptualization of Personal Autonomy" (Ahlin Marceta 2018), I develop an account of personal autonomy which includes the notion of authenticity. I apply it in an analysis of a paradigm case of possible inauthenticity to test and demonstrate the practical usefulness of the conceptualization. I call this article "the Autonomy article." Finally, the fourth article, entitled "Nine Cases of Possible Inauthenticity in Biomedical Contexts and What They Require from Bioethicists," collects various cases in biomedical contexts where the notion of authenticity has been or could reasonably be expected to be of moral significance. The account developed in the Autonomy article is applied to the cases where this is possible. In what follows, I call the article "the Nine Cases article," or variations on that theme.

The introduction is structured as follows. In the next section, I provide an overview of the most central concepts that are relevant for the present purposes. Thereafter, I give a detailed description of the main contribution of this thesis, namely that it bridges the gap between theoretical ideals of authenticity and authenticity-related problems that clinicians face in practical biomedical contexts. The subsequent section includes a methodological discussion of how I have approached this problem. Among other things, reflective equilibrium as a theory of moral justification is explained and the methodological choices of this thesis are spelled out. In the section thereafter, I summarize the four articles. In that section, I also discuss some views that I have had to revise since the publication of the articles, defend some of the choices I have made with regard to the present purposes, elaborate on the theoretical and practical context of this thesis, and show how the articles are connected to each other.

Central concepts

For the present purposes, the most central concepts are *personal autonomy*, *decision-making capacity*, and *voluntariness*. This section, parts of which have been published in Ahlin (2017), provides a brief overview of the three concepts. The discussion also places the arguments in this thesis into a conceptual context.

Personal autonomy

There is no consensus regarding how personal autonomy should be understood. But, the "many faces of autonomy" may not be as numerable as some have suggested (Taylor 2009, Ch. 2). It is generally held that autonomy, in the moral sense relevant to the present discussion, is a property that can be enjoyed to different degrees. As a matter of degree, autonomy is not a binary concept; a person can be *more* or *less* autonomous, as well as *not* autonomous and *fully* autonomous. Furthermore, autonomy is a property with both positive and negative elements. Positively, autonomous persons are, for instance, capable of qualitative self-reflection; they can assess their own desires and values and choose whether to be moved by them. Negatively, autonomous persons are not subject to control by other agents, influences, or conditions.

In contemporary theory, the distinction should be made between *procedural* and *substantial* accounts of personal autonomy. In the procedural tradition, autonomy only concerns the *form* that decisions and actions take. Theorists are here only interested in matters such as the process by which an agent comes to make a decision, the independence of her choosing relative to external influences, and so on.

In the substantial tradition, autonomy also concerns the *content* of decisions and actions. In addition to matters of a procedural nature, some substantialists take an interest in whether an agent's choices are self-supporting. To exemplify, consider a person who is physically and mentally abused by her partner. The victim reflects upon whether to leave her partner, but decides not to do so. When contemplating the case, proceduralists take into consideration the process by which the

victim makes her decision, putting weight on the independence of her decision-making procedure. They may conclude that the victim made an autonomous decision. Substantialists, on the other hand, are concerned also with the fact that the victim chose not to leave her abusive partner. They may instead conclude that the victim's choice is self-injurious rather than self-supporting, and that it is therefore non-autonomous.

Proceduralists sometimes accuse substantialists for unjustified paternalism, to which the latter tend to reply that proceduralists unwarrantedly ignore the social embeddedness of personhood. I will not engage with that debate here. In what follows, I will only treat issues in the procedural tradition, in line with the standard accounts in medical ethics.[1]

There are three major ways in which personal autonomy is relevant for the present purposes, namely autonomous *wishes*, *decisions*, and *acts*. The autonomy of wishes and decisions concerns the inner life of agents while the autonomy of acts concerns their outer life. A person can, for instance, hold autonomous wishes and make autonomous decisions, but for some reason be incapable of autonomously acting upon those wishes and decisions. To illustrate, consider a fully healthy patient who is strapped to her hospital bed due to a clinician's mistaken belief that she will hurt herself and others if left unconstrained. The patient is unable to move freely, and is thus robbed of her capacity to act autonomously. Yet, she can hold the autonomous wish to be freed, and make the autonomous decision to try to free herself by twisting and turning violently to break out of the straps.

Likewise, a person can be capable of acting autonomously while holding non-autonomous wishes and making non-autonomous decisions. Consider a patient who is temporarily under the influence of drugs that do not affect her physical abilities but significantly distorts her view of herself and her surroundings. She might, for instance, hold a non-autonomous wish to hurt herself, non-autonomously decide to do so, and autonomously act upon those wishes and decisions.

[1] For further inquiry into the debate between proceduralists and substantialists, see, e.g., Christman (2004, 2015) and Oshana (2015).

In their book *Principles of Biomedical Ethics* (2013), Tom L. Beauchamp and James F. Childress hold respect for autonomy as one of four principles that in combination encompasses biomedical ethics. The other principles are *nonmaleficence*, i.e., the obligation to abstain from causing harm to others, *beneficence*, i.e., the moral requirement to contribute to others' welfare, and *justice*, i.e., equality in access to health care and in health status. In the book, none of the four principles take precedence over the others *a priori* (cf. pp. 13–25). However, according to some bioethicists, respect for autonomy is "first among equals" (Gillon 2003, p. 310):

> Firstly, autonomy—by which in summary I simply mean deliberated self rule; the ability and tendency to think for oneself, to make decisions for that thinking, and then to enact those decisions—is what makes morality—any sort of morality—possible.

In what follows, I do not commit to any particular position regarding the moral weight of personal autonomy, beyond the general recognition that it is morally valuable in biomedical contexts.

A more detailed account of personal autonomy is introduced in the Autonomy article below. To summarize the discussion in this subsection, autonomy is a property that persons can enjoy to different degrees. In this context, it matters to a patient's autonomy whether she is capable of making healthcare decisions. Lacking such competence entails that she is non-autonomous in some aspects and to some extent. Furthermore, it matters to a patient's autonomy whether she makes her healthcare decisions voluntarily; non-voluntary decision-making is non-autonomous. These concepts are explained in greater detail in the two following subsections.

Decision-making capacity

In these contexts, competence is an element that refers to a patient's capacity to make healthcare-related decisions. A patient is competent, or has decision-making capacity, if she can understand information provided, appreciate in what way it concerns her, and reason about it in light of her own values and preferences (cf. Charland 2015, sec 2). These

capabilities imply several others. For instance, they require of patients that they are capable of thinking critically of themselves as intertemporal subjects; a capability often lacking in children (and others). Beauchamp and Childress suggest seven types of related inabilities (2013, p. 118):

1. Inability to express or communicate a preference or choice
2. Inability to understand one's situation and its consequences
3. Inability to understand relevant information
4. Inability to give a reason
5. Inability to give a rational reason (although some supporting reasons may be given)
6. Inability to give risk/benefit-related reasons (although some rational supporting reasons may be given)
7. Inability to reach a reasonable decision (as judged, for example, by a reasonable person standard)

These mark a threshold level of decision-making competence, so that persons who display one or more inabilities from 1 through 7 should be judged as not fully competent to make the decision in question.

Competence is not a *global* but a *particular* threshold element, in the sense that being competent is to be competent relative to some specific decision (Buchanan and Brock 1990, pp. 18–20). For example, a person may be capable to make decisions about her healthcare but not about her finances, or capable to make one healthcare decision in the morning but incapable to make that same decision in the evening.

Furthermore, Beauchamp and Childress recognize that the level of evidence for determining competence should vary according to the risk of the decision; complex health care decisions should require a higher degree of confidence in the patient's decision-making competence than simple decisions. For instance, the required level of evidence of competence should be higher when the decision is to consent to participation in medical research than the required level of evidence when the decision is to object to participation (2013, p. 120). Therefore, it must be determined in each case against objective standards whether a patient is competent relative to the particular decision in question.

In an article which is cited below, namely Grisso et al. (1997), an instrument is presented for assessing patients' decision-making capacities in clinical practices. The instrument is called "the MacArthur Competence Assessment Tool—Treatment" (MacCat-T). Applied in interviews with patients, the MacCat-T tests abilities related to understanding of relevant information, reasoning about the risks and benefits of potential options, appreciation of the nature of one's situation and the consequences of one's choices, and to expressing a choice (p. 1415).

An interview requires 15 to 20 minutes. During this time, *understanding* is assessed by evaluating the patient's ability to paraphrase what has been disclosed concerning her disorder, the recommended treatment, and related benefits and risks (p. 1416). *Reasoning* is assessed by examining how the patient explains her choices, i.e., whether she mentions relevant consequences, alternatives, etc., and if her choices are coherent with her explanation of them (ibid). *Appreciation* is assessed by exploring if the patient acknowledges that the relevant information applies to her; lacking appreciation is shown if the patient's beliefs are based on delusional or distorted perceptions (ibid).

Voluntariness

I call the theory of voluntariness which has been most influential in bioethics *the voluntariness-as-control theory*. It is supported by, among others, Appelbaum et al. (2009), Beauchamp and Childress (2013), and Nelson et al. (2011). According to the voluntariness-as-control theory, an action is voluntary if it is free from controlling influences.

Nelson et al. (2011) provide the most elaborate account of the theory that voluntariness is closely linked to being in control over one's actions.[2] Voluntary action, they argue, should be understood in terms of the two necessary and jointly sufficient conditions of *intentional action* and

[2] According to Nelson et al. (2011, p. 11), the theory of voluntariness as degree of control was first introduced by Wall (2001). However, Wall did not conceive the notion of voluntariness as control. Beauchamp and Childress had already written that the "primary meaning of 'voluntariness' is exercising choice free of coercion or other forms of controlling influence by other persons" in the second edition of their *Principles* (1983, p. 87).

absence of controlling influences (p. 6). The notion of intention is binary, in the sense that an act either is or is not intentional, while the notion of controlling influences is a matter of degree, so that an act can be more or less free from controlling influences on a continuum from total control to total absence of control.

Examples of controlling influences in the broad sense include offers of payment, threats, education, deceit, manipulative advertising, emotional appeals, and the like (p. 7). Such influences can deprive agents of at least some degree of voluntariness. Manipulation involves "the use of nonpersuasive means to alter a person's understanding of a situation and motivate the person to do what the agent of influence intends" (p. 8). A person can be manipulated in several ways. One can manipulate the information a person receives through different communication techniques or the format and method of risk disclosure. Financial incentives such as offers or rewards or access to drugs or medical care can distort a person's view of her options of choice.

Furthermore, one can be manipulated through, for example, withheld information, misleading exaggeration, and explicit lies, which are all examples of cases in which the manipulated agent has no credible possibility of recognizing that she is receiving skewed information. Similarly, a person may be persuaded into doing or believing something. However, Nelson et al. argue that persuasion is consistent with voluntariness. When persuaded, "a person believes something through the merit of reasons proposed" and is therefore not controlled (p. 7).

Finally, a person can be controlled through coercion. Building from a conceptual framework that was first introduced by Nozick (1969), Nelson et al. conceptualize coercion as the total control over an agent's actions that occurs "if and only if one person intentionally either forces another person or uses a credible and severe threat of harm to control another person" (p. 7). True coercion by threat "requires that a credible and intended threat disrupts and reorders a person's self-directed course of action" (p. 8).

It has been suggested that voluntariness presupposes authenticity. More specifically, the proposal is that voluntary choice requires choosing

"in a way that is in conformity with one's identity, affective state, values, and goals, and is truthful to one's sense of self and view of the good life" (Berghmans 2011, p. 24). I am sympathetic to the moral idea of analyzing autonomous choice in terms of authenticity, but for reasons of analytical clarity and precision I think that it is better to treat authenticity as an independent concept rather than to include it in the theoretical base of the conceptualization of voluntariness.

The aim of this thesis

To repeat, the aim of this thesis is to make theoretical ideals of authenticity helpful in practical biomedical contexts, with focus on a theoretical ideal of authenticity known mainly from Harry G. Frankfurt (1971) and Gerald Dworkin (1988) and on practical problems concerning medical decision-making. This aim is facilitated by an overview of authenticity-related problems (the Nine Cases article) and an explanation of why theoretical ideals of authenticity are unhelpful in practice (the Determining Authenticity article). I argue that the aim is attained in two respects; the thesis further develops an already established theory of authenticity so that it yields practically observable implications (the Inauthenticity Judgments article) and proposes an authenticity-based conceptualization of personal autonomy against the backdrop of those implications (the Autonomy article).

The current authenticity-related moral problem

The Nine Cases article begins with a quotation from a person who reports of her anorexia nervosa: "I wasn't really bothered about dying, as long as I died thin" (Tan et al. 2006, p. 274). Anorexia sometimes affects how people who suffer from it value themselves, i.e., mainly their weight and body size, and in turn the values affect the anorectics' motivational sets with regard to nutrition and care. Thus, there is sometimes a problematic interaction in play between the disorder and the values that anorectics have. In some cases, anorexia nervosa patients have decision-making

capacity. Yet, they hold values that seem problematic in the above sense; some report that they would rather die than gain weight.

Intuitively, there is something deeply distressing about holding such values, and this distress has led some to analyze cases of anorexia nervosa in terms of authenticity (Hope et al. 2011; Sjöstrand and Juth 2014; Tan et al. 2006). One suggestion is that while some anorectics may have decisional capacity, they are in a state of inauthenticity (ibid). That is, they are not themselves, in some substantive sense, and should therefore nevertheless not be allowed to make their own healthcare decisions. Or, that is the hypothesis which motivates the aim of this thesis.

Similar problems also appear in other medical situations. Untreated syphilis may cause changes in a person's character that make the person or her decisions inauthentic. People suffering from borderline personality disorder (BPD) may, in a short time span, express drastically conflicting opinions on their medical treatment. It may be the case that their conditions should be described and analyzed in terms of inauthenticity. And so on. The Nine Cases article collects nine examples in which the notion of authenticity has been or may be relevant in practical biomedical contexts.

Philosophers that have set out to analyze authenticity in biomedical contexts have proposed various conceptualizations of the notion. There is substantial disagreement already at the outset of this debate. First, it is not clear what it is that should be subject to critical scrutiny in terms of authenticity. Some hold that an analysis of authenticity must begin with the concept of what it is to be a real person. Others hold that the notion of personhood is secondary at best, as it is the authenticity of medical decisions that is of interest in clinical practices. Secondly, philosophers who agree on what should be the subject of the analysis champion competing theories of what distinguishes authenticity from inauthenticity. For instance, some theorists argue that it is the causal history of a desire that matters most to its authenticity. Others, while agreeing on the focus on desires, instead argue that it is the coherence of full desire-sets that matters.

Choices and delimitations

In this thesis, I have made two choices with regard to these debates, neither of which will be defended at length. The first choice is to focus on the authenticity of desires, rather than of persons, lives, or something else. This is because I, as many others in this field, hold desires to be the most basic element in ordinary preference forming and, thus, the most basic element in decision-making (cf., e.g., Taylor 2005). For the purposes of this thesis, I think of a desire as an attitude or directedness which influences the decisions that the desire-holder makes. Among other things, this means that some of the problems that are introduced in the Nine Cases article are not treated in this thesis, as they do not concern decision-making.

The second choice I have made is to focus on a theoretical tradition of thinking about authenticity that first took form in a set of books and articles in the 1970's and 1980's, of which Frankfurt (1971) and Dworkin (1988) are the most noteworthy. In this tradition, authentic desires are distinguished from inauthentic desires in that the former would be endorsed, at least hypothetically, by the desire-holder upon informed and critical self-reflection. Here, I call this criterion "affirmative self-reflection."

I have had two reasons for making this choice. First, the Frankfurt–Dworkean tradition of thinking about authenticity in terms of second-order volition, meaning that the distinguishing feature between authenticity and inauthenticity lies in the agent's self-perception of the desire in question, has been more influential than any other theoretical tradition with its roots in the last four or five decades of bioethical inquiry. If there is one mainstream theory of authenticity, this is it. It is well-known to my intended audience and any contribution to it should be of interest to autonomy theorists in general, and to authenticity theorists in particular. Thus, it is a reasonable choice to attempt to contribute to this theoretical tradition rather than to some other, more peripheral, tradition.

Secondly, I think that theories in this tradition do a better job in distinguishing between authenticity and inauthenticity than other types of theories. It is an intuitive understanding that "authenticity" is a property of a person's desires that makes them different from desires that she does

not want to have. The most basic reason to consider the "authenticity" of desires is that we want to distinguish between such different kinds of desires. The Frankfurt–Dworkean tradition of thinking about authenticity manages to make this intuitive understanding of authenticity theoretically plausible. It provides an explanation of authenticity that, in light of the arguments from its advocates, seems very reasonable.

The problem of practical application

There are various problems with this theory. For instance, it can be argued that affirmative self-reflection itself requires affirmative self-reflection, and that the theory therefore results in an infinite regress.[3] I do not address such problems in this thesis, i.e., problems concerning whether the theory succeeds in distinguishing between authenticity and inauthenticity. At present, I do not have a more elaborate defense of this tradition than what other theorists have already put forth (see, e.g., Christman 2009 and Juth 2005). However, it is not included in the aim of this thesis to defend this kind of theory as such. In what follows, my arguments should be understood as building on the assumption that some version of the theory is true, or at least plausible.

My focus is instead on problems associated with applying the theory in practical contexts. One major problem is that the theory fails to yield practically observable consequences. I elaborate on this in the Determining Authenticity article. In short, it is difficult to know whether a desire-holder would endorse her own desires upon informed and critical self-reflection. Therefore, affirmative self-reflection appears to be an ideal that is unhelpful in terms of action-guidance in practical biomedical contexts.

In the Inauthenticity Judgments article, I develop a version of the Frankfurt–Dworkean theory that includes practically observable indicators of inauthenticity. This is a first attempt to bridge the gap between theoretical ideals of authenticity and practical authenticity-related prob-

[3] See Taylor (2005) for an elaborate version of this argument. See also Juth (2005, pp. 153–62) for a version of the theory in which the problem of an infinite regress may never arise.

lems. The version of the theory is not morally neutral. It is formulated in terms of moral justification, meaning that the problem of determining whether a desire is inauthentic is phrased in terms of when it is morally justified to make the judgment that it is inauthentic. Among other things, this means that there may be inauthentic desires that observers are not justified to judge as inauthentic.

It is important to note that paternalist interventions such as, for instance, force-feeding an anorectic who states that she would rather die than gain weight, are not justified merely because it is (by hypothesis) justified to judge that her desires are inauthentic. Paternalist interventions require further justifications, not least considering the proportionality of the intervention and the degree of epistemic certainty of inauthenticity. This thesis does not include any elaborate discussion of paternalism, or any detailed suggestions of how the present theories could support paternalist interventions. With that being said, the background of the discussion is the practical bioethical problem of compulsory care, and whether there is any ground for using considerations of inauthenticity as part of the justificatory base for overriding someone's healthcare decisions.

In the Autonomy article, I incorporate my re-stated and morally loaded version of the Frankfurt–Dworkean theory in Beauchamp and Childress's theory of personal autonomy. Thereby, this thesis constructs a conceptualization of autonomy that manages to take authenticity into account. The principle of respect for autonomy is widened to include judgments of authenticity, which is one way in which this thesis makes theoretical ideals of authenticity helpful in practical biomedical contexts. In my view, this thesis contributes to solving a paradigm problem which has concerned theorists and practitioners for some time. Although it is of course up to the critical reader to judge, I believe that the aim of the thesis has thus been met.

Methodological issues

In this section, I discuss a number of methodological issues connected to the present purposes. First, I explain my initial approach to the problems treated in this thesis. Then, I introduce reflective equilibrium as a theory

of justification and as a process of moral inquiry. Thereafter, I discuss the kind of normative guidance that theorists in this context can provide practitioners with. A final subsection includes some summarizing concluding remarks.

Initial approach to the problem

The first methodological choice in applied ethics is how to approach the problem one wishes to solve. There are various such possibilities. Sometimes, theorists begin by distinguishing between top-down and bottom-up approaches. In other cases, they begin by distinguishing between ideal and non-ideal theory.

In a top-down approach, a theory is chosen and applied to the problem at hand. This is how, for instance, Peter Singer approaches problems in practical ethics (Singer 1993). In his book, Singer takes a "broadly utilitarian position" on various moral topics, such as animal rights, abortion, and the environment (p. 12). Thereby, he attempts to solve practical problems by applying utilitarianism to them and report of the results. One problem with the top-down approach is that it demands a lot from the theory that is applied. Singer's proposed solutions are dependent on the truth of utilitarianism, which is far from evident.

In a bottom-up approach, the goal is to identify the (potential) problem independently from normative theories first, and then apply different theories to see what comes out of the analysis. This is, for instance, how Jonathan Wolff approaches moral topics in public policy such as gambling, drugs, and safety (Wolff 2011). In his book, Wolff seeks to describe the cases neutrally, before engaging in moral analysis from the perspective of different normative theories (ibid, "Introduction"). One problem with the bottom-up approach is that moral problems cannot be observed with no prior idea of what is morally relevant. The mere fact that something is described as a moral problem appears to signal that some tacit normative assumptions have been made.

Sometimes, more often in political philosophy than in applied ethics, theorists instead begin by distinguishing between ideal and non-ideal

theory. There are different uses of the terms (see, e.g., Hamlin and Stemplowska 2012; Valentini 2012), but I understand them as follows.

In ideal theory, theorists construct a desirable hypothetical model of the object under scrutiny. Thereafter, the actual circumstances are compared to the model, and action-guidance is provided through observation of the differences. This is, for instance, how John Rawls (2001) takes on the problem of justice. In his writings, Rawls constructs a hypothetical model of society in which its basic institutions are perfectly just. Real societies can be compared to the Rawlsian ideal and policies may be formulated which would lead society in the direction of the ideal, or perhaps even fulfill it.

It can be argued that one problem with ideal theory is that although hypothetical models may be very neat, real people and real institutions rarely behave as expected. Ideals do not take the complexity and imperfections of the real world into sufficient account and therefore sometimes, or in some respects, fail to provide substantial guidance for vacillating agents (see, e.g., Sen 2006; Wiens 2015).

Non-ideal theory, as I will understand it here, is more similar to the bottom-up approach. It is problem-oriented, in the sense that a problem is identified, the possibilities and limitations of the case are explicated, and proposals are formulated as to what might make the case less of a problem. This is, for instance, how David Schmidtz takes on the problem of justice (2006). In his book, Schmidtz proposes that the notion of justice has four elements which are expressed by principles of *desert*, *reciprocity*, *equality*, and *need*. For any given justice-related problem, principles from the four elements should be weighed against each other to articulate and solve the particular problem at hand.

Defenders of ideal theory have argued, among other things, that idealization is nonetheless a necessary component in moral thinking. For instance, O'Neill writes that "if ethical principles are to be relevant to a wide range of situations or of agents, they surely not merely *may* but *must* be abstract" (1987, p. 55). Much of the criticism of ideal theory, it has been argued, is "too sweeping" (Erman and Möller 2013, p. 43).

In this thesis, I have attempted a moderately problem-oriented approach, in the sense that I have taken on what appears to be moral problems (and have been treated as moral problems by other ethicists). I have not approached them top-down or constructed ideal models in my attempts to provide normative guidance with regard to them. Yet, I do not claim to approach the problems from a fully neutral point of view. Most importantly, I think of my view as guided by the individualist, autonomy-based, non-paternalist contemporary bioethical paradigm (Ahlin 2017; Faden and Beauchamp 1986; Jonsen 2000). Furthermore, I am methodologically guided by a theory of what justifies normative propositions, namely reflective equilibrium. Among other things, this means that I take on issues that are intuitively problematic, aiming to provide a balanced normative judgment with regard to them. Thus, my initial approach to the problems in this thesis is non-ideal, mainstream, and methodologically theory-dependent.

Normative justification

I adhere to a theory of what justifies normative propositions, and of how moral inquiry should be conducted, that is commonly known as reflective equilibrium. It is a coherentist theory that bases justification on the coherence of a full set of beliefs, to be contrasted with, e.g., foundationalist theories in which justification rests on a non-inferential foundation (Daniels 2016; Hasan and Fumerton 2016). In this subsection, I elaborate on my view of reflective equilibrium as a theory of justification and as a process of deliberation, beginning with the former.

For the present purposes, I distinguish between claims of knowledge, truth, and fact on the one hand, and claims of justification on the other. As a theory of justification, reflective equilibrium is a theory about reason-giving (cf. de Maagt 2017, pp. 445–6). In short, I take it to be the theory that a normative claim is reason-giving to the extent that it is coherent with all other beliefs (in moral and non-moral matters) and with our stable and considered moral intuitions. For instance, the claim, "*ceteris paribus*, patients who suffer should be helped" is justified not

because it is true, but because it is coherent with all other beliefs and with our moral intuitions.

Among other things, moral reasons vary in strength and in relevance. The strength of a moral reason is relative to other elements in the equilibrium, not relative to some independent scale of measurement. For instance, relieving suffering is sometimes a reason to intervene with a patient's healthcare decisions, but respecting the patient's autonomy is often a stronger reason not to intervene. Promoting justice between patients is also a moral reason, although it is not relevant for the present discussion. A judgment on whether to intervene with the patient's healthcare decisions is justified to the extent that it is balanced, taking all relevant moral reasons, beliefs, and intuitions in consideration.

So far, reflective equilibrium has been treated as a hypothetical end-state of a deliberative process in which normative claims are justified. But, it is common for the term to designate the deliberative process itself, i.e., a method of moral inquiry.

As a method, reflective equilibrium is the deliberative process of reflecting on and revising moral judgments (and judgments related to them). In that process, empirical facts, risks and uncertainties, critical self-reflections of possible biases and other cognitive misbehaviors, and so on, must be taken into account. The process is goal-driven. It aims to identify what is (and what is not) reason-giving in a particular case and provide a balanced and considered judgment with regard to it. Furthermore, my view is that the process should be thought of as continuous, in the sense that theorists should treat moral justification as an ongoing process of evaluation and re-evaluation where progress is made successively.

There are various criticisms of reflective equilibrium.[4] Perhaps most commonly, it is argued that for any subject there may be two (or more) internally coherent sets of reflective equilibria, meaning that the method cannot provide conclusive normative guidance. Another common criticism is that there is no guarantee that any particular reflective equilibrium is not in fact merely a coherent set of ungrounded

[4] See, e.g., Daniels (2016) for a general overview and, e.g., Strong (2010) and Willigenburg (2007) for criticism that is specific for reflective equilibrium in biomedical contexts.

prejudices. Although these criticisms should be taken seriously by bioethicists, I will not elaborate on them here nor provide a detailed defense of reflective equilibrium.

Theorists and practitioners

The aim of applied ethics is sometimes to suggest practical policies, such as how to distribute scarce healthcare resources or what an informed consent form should contain. At present, however, my aim is rather to contribute to a framework for decision-making. The framework is constructed so that it spells out some normative content, such as which moral principles should be respected, but leaves some blank spaces which must be filled in by practitioners, such as what respecting one of the moral principles entails in a particular case. Therefore, the framework includes "instructions" for "users."

As explained above, respect for autonomy is a central moral principle in bioethics. Among other things, the principle obliges healthcare practitioners to refrain from intervening with patients' decisions concerning their own healthcare (possible exceptions include when patients lack decision-making capabilities or are subject to controlling influences). Theorists can spell out in greater detail what respect for autonomy entails generally, but it is difficult to formulate precisely how the principle applies in particular cases. For illustration, consider this example, which builds from Lee (2010, p. 525).[5]

> A 17-year-old has lost a lot of blood in an accident. The best chance of saving the teenager's life is an urged blood transfusion and a surgical intervention to stop the bleeding. However, the teenager's parents are Jehovah's Witnesses. For religious reasons, they refuse to give permission for the blood transfusion. They request that surgery should be carried out anyway, although they understand that this will be much more dangerous than operating with blood transfusion.

[5] In Lee's original example, the case concerns a 2-year-old child.

In this case, there is a conflict between the principles of beneficence, i.e., to do good, and respect for (surrogate) autonomy. One considered judgment is that the principle of respect for (surrogate) autonomy should be overridden for beneficence-related reasons and that the doctor should proceed with transfusion of blood to the teenager.

This judgment is not anticipated by bioethical theory but is the result partly of theory and partly of practical judgment in the particular case. That is, nowhere it is written or in any other way stated what should be done in cases which involve 17-year-olds who have lost a lot of blood in accidents and the best chance of saving the teenagers' lives is through urgent blood transfusions and surgical interventions to stop the bleeding, and parents who are Jehovah's Witnesses and refuse to give permission, and so on. There are no indexes that include every conceivable bioethical dilemma that practitioners can consult in search of moral guidance. One way to phrase this indeterminacy is that the moral principles involved are underdetermined (O'Neill 1987).

Bioethicists in the current school of thought inform practitioners about how they should make moral decisions in cases where there is little or no pre-existing guidance. This has two implications of relevance to the present purposes. First, some normative content is determined in practical settings rather than in theory. Healthcare practitioners, who are in direct contact with moral dilemmas, must be expected to be well-equipped and trained moral decision-makers. Second, there are methods for applying underdetermined bioethical principles. I will briefly treat two such methods below, namely specificationism and casuistry.[6]

Methods for moral decision-making

When applying abstract and underdetermined moral principles and concepts, practitioners determine some of their normative content. That can be done better or worse; better if it is done methodically, and worse if not. Specificationism and casuistry are two methods for this kind of moral

[6] It should be noted that it has been argued that there are no real differences between specificationism and casuistry. See, e.g., Cudney (2014) for an illuminating discussion.

decision-making. On my understanding, both aim at providing reliable moral justification within a framework of reflective equilibrium, but in different ways. With the exception of one section in the Autonomy article, the present thesis does not include practical applications of abstract and underdetermined moral principles and concepts. Therefore, the discussion in this subsection should be understood as being forward-looking; it briefly introduces the methodological basis of how the normative substance in this thesis should be applied in practical settings.

Following Rauprich (2011), the first step in both specificationism and casuistry is to decide tentatively which moral principles that apply in the case at hand. For instance, in the case cited above with the 17-year-old whose parents refused a blood transfusion, little deliberation is required to determine tentatively that there is a conflict between beneficence and respect for (surrogate) autonomy. The differences between the two methods begin to appear in the second step of the process, which aims to determine how the relevant moral principles apply in the particular case.

In specificationism, the second step is interpretative. One interpretation of the principle of beneficence is that "it is morally prohibited to risk the death of a patient if his or her life-threatening condition can be medically managed by suitable medical techniques," and an interpretation of the principle of respect for (surrogate) autonomy is that "it is morally prohibited to disrespect a parental refusal of treatment" (Lee 2010, p. 525). A balanced judgment may be that "it is morally prohibited to disrespect a parental refusal of treatment unless the refusal constitutes child abuse or child neglect or violates a right of the child," and that the parents' refusal does in fact constitute abuse, neglect, or a rights-violation (ibid, pp. 525–6). Interpretation requires insight into the content and purpose of the moral principles, and an understanding of the relevant empirical facts associated with the case at hand. It also requires an explanation of why the chosen interpretation is correct.

In casuistry, the second step is comparative; guidance is sought in comparisons with similar cases. For instance, in one similar (hypothetical) case a decisionally-incapable adult is offered vaccination against Hepatitis A, which the surrogate decision-maker refuses with reference

to the irrational and uninformed belief that vaccines cause autism. Given that the cases are sufficiently similar, the comparison provides guidance in the case at hand. Suppose that the practitioners in the hypothetical case decided to override the surrogate decision-maker's wish. Then, the practitioners in the case with the 17-year-old have reason to decide to proceed with the blood transfusion. Among other things, casuistry requires evidence of the similarity of the cases being compared. The cases that are used for comparison should preferably be paradigm cases, i.e., cases in which it is reasonably clear what should be done (Strong 2000, p. 331). But, they may also be hypothetical. Then, they are thought examples of the kind that is common to the ordinary philosophical method of principled argumentation.

Both specificationism and casuistry can be further elaborated (see, e.g., Beauchamp and Rauprich 2016 and Strong 2000). However, I will not provide more detailed accounts of the two methods. The brief introduction above suffices for the present purposes, i.e., to give a general idea of how the arguments in this thesis can contribute to all-things-considered judgments about how to act in particular situations in healthcare.

Concluding remarks on methodological issues

In conclusion, this thesis aims to make theoretical ideals of authenticity helpful in practical biomedical contexts. I approach this problem from a non-ideal yet theory-dependent point of view. Most importantly, I adhere to reflective equilibrium as a theory of justification and as a method of moral inquiry. I recognize that in this context, normative principles and concepts are underdetermined, i.e., that some normative content is determined in practical settings rather than in theory. Therefore, practitioners should be equipped with and trained in the methods of reliable moral decision-making. I have here briefly mentioned two such methods, namely specificationism and casuistry. It is beyond the purposes of this thesis and the methodological issues that accompany its aims to elaborate further on these matters here.

Summary and discussion of the articles

Article 1: The Determining Authenticity article

In this article, I develop a taxonomy of characteristics displayed by various theories of authenticity that enables overview and analysis. Thereafter, I use the taxonomy to argue that no category or class of characteristics yields practically observable consequences. I conclude that in practice, the authenticity of desires cannot be reliably determined, and that authenticity should therefore not be employed in informed consent practices in healthcare. Since the publication of this paper, I have had to revise some of the views expressed in it. For instance, as the aim of this thesis suggests, I am no longer of the view that authenticity has no role to play in informed consent practices. Therefore, in addition to summarizing this article, I will here also explain in detail my current views on the central topics discussed in it.

The article takes as its starting point the concept of informed consent, which denotes a patient's valid consent to or refusal of a medical intervention. In simple terms, informed consent is short for informed, voluntary, and competent consent (Eyal 2012). The general understanding in bioethics is that informed consent aims to protect and promote patients' autonomy, although alternative interpretations have been suggested (see, e.g., O'Neill 2003). The problem treated in the article is whether authenticity should be among the conditions of informed consent.

To give the problem a practical context, I introduce the hypothetical case of Anna, a professional ballet dancer who needs medical treatment. Anna is informed about her situation, is competent to make healthcare decisions, and does so voluntarily. Yet, she makes the surprising decision to refuse to undergo a treatment that would allow her to continue dancing. Her doctor considers whether Anna's decision is authentic, and whether her "true wishes" could be adhered to by forcing her to undergo the medical procedure; perhaps Anna's refusal is invalid.

Then, I introduce what I call "the argument from testability." The argument from testability is that worries such as the one that Anna's doctor has not only require a theory of authenticity but also the ap-

propriate means to *test* the authenticity of patients' decision-making, which is difficult. It first appeared in Sjöstrand and Juth (2014) and is foundational to this thesis; essentially, the aim of making theoretical ideals of authenticity useful in practical contexts includes making it practically feasible to test the authenticity of decision-making.

One of the merits of the Determining Authenticity article is that it stresses the significance of the argument from testability by elaborating on it and applying it to various traditions of authenticity theory. However, in the concluding remarks of the article I claim to have shown that the authenticity (or inauthenticity) of desires cannot be reliably detected. That is an overstatement. I no longer believe that there is support in the article for the claim.

In the article, instead of going through various theories of authenticity and analyze them individually, I attempted to support the claim that authenticity cannot be reliably detected by generalizing features that various authenticity theories share and examine them categorically. This led to a taxonomy of features that various authenticity theories share. Although the purpose of developing the taxonomy was methodological, i.e., it was introduced for evaluative purposes, the taxonomy itself is another merit of the article. It provides a systemic overview of authenticity theories that enables analysis. In combination with a similar taxonomy by Robert Noggle (2005), I use the taxonomy again—for other purposes—in the Inauthenticity Judgments article.

According to the taxonomy, the features that authenticity theories share can be organized into three categories, namely *sanctionist*, *originist*, and *coherentist*. In sanctionist theories, i.e., theories that display features from the sanctionist category, desires are authentic if they are endorsed by the desire-holder upon self-reflection. The Frankfurt–Dworkean tradition of thinking about authenticity belongs to this category, and in what follows I will only use the categorical term "sanctionism" to denote it. In originist theories, desires are authentic if they have the right kind of origin. In coherentist theories, desires are authentic if they are coherent with the desire-holder's full set of desires. Furthermore, the features are organized into two classes, namely *cognitivist* and *non-cognitivist*.

In cognitivist theories, authenticity is a matter of rational deliberation. Non-cognitivist theories do not commit to this view.

In the article, I claim that the taxonomy is exhaustive, with the exception that it does not include theories of authenticity from the substantial tradition of autonomy theory. However, it should be noted that the taxonomy does not cover theories of authenticity in an existentialist tradition either, i.e., theories that may be found in, e.g., Heidegger or Sartre. This thesis is only concerned with theories of authenticity from the procedural tradition, in which authenticity is analyzed according to the content-neutral processes by which desires are formed or are sustained. To be clear, I am now of the view that the taxonomy only covers theories from this tradition.

After having introduced and explained the taxonomy, I discuss what the argument from testability requires from each category. This is where the significance of the argument from testability is highlighted, although the arguments in the article do not support the claim that no theory passes the test. I return to this discussion in both the Inauthenticity Judgments article and the Autonomy article. In those articles, I build on the weaker view that while it may not be impossible to reliably determine the authenticity of desires, it is nonetheless difficult to do so.

In the concluding remarks, I claim that authenticity should not be included as a criterion in informed consent. Although I am currently ambivalent about whether authenticity should be among the conditions for valid consent or refusal to medical interventions, I am certain that the arguments in the article do not suffice to ground the claim. In the Autonomy article, I argue that autonomous actions and choices may be analyzed in terms of authenticity. Among other things, this may enable the inclusion of authenticity in informed consent practices, but the possibility is not discussed further in this thesis.

Article 2: The Inauthenticity Judgments article

In this article, I argue that under certain conditions it is justified to judge that a desire is inauthentic. My argument is threefold. First, I propose a sanctionist thesis of the conditions under which judgments

of inauthenticity are justified. Then, I introduce two empirical factors that, when combined, indicate that the conditions in the thesis are met. Finally, I delimit the scope of my arguments to target only a certain kind of people and a certain kind of desires. Since the publication of the article, I have had to revise my views about the final delimiting clause. As in the previous subsection, I will here not only summarize the article but also introduce my current views with regard to the arguments in it.

The sanctionist thesis which I propose in the article is that judgments of inauthenticity are justified if there is sufficient reason to believe that the desire-holder would disapprove of having the desire upon informed and critical self-reflection. I call it "the dissenting self-reflection thesis" to connect it to, but also distinguish it from, the sanctionist ideal of affirmative self-reflection. This reversed version of the sanctionist ideal is a re-statement of the central thesis in sanctionism as a moral thesis; the dissenting self-reflection thesis does not distinguish between authenticity and inauthenticity, but between when it is justified to judge that a desire is inauthentic and when it is not justified to do so. Among other things, the dissenting self-reflection thesis entails that there may be inauthentic desires which, mainly for reasons of epistemic uncertainty, it is not justified to judge as inauthentic.

After having introduced the dissenting self-reflection thesis, I suggest two empirical factors which in combination would indicate that a desire-holder would disapprove of having a desire, i.e., that it may be justified to judge that the desire is inauthentic. The first indicator of inauthenticity is if it is known that the desire is due to causal factors that are not normal to how the desire-holder is otherwise construed, taking both physical and mental dispositions into consideration. The second indicator is if it is known that the desire does not cohere with how the desire-holder's identity has developed over time and is presently being sustained. In the article, the two indicators are scrutinized. It is shown why both must be present for judgments of inauthenticity to be justified.

However, the dialectic in the article only leads to intuitively sound conclusions regarding desires that are bad, in some sense. When a desire is good, in some sense, it does not seem to be justified to judge that it is

inauthentic in spite of the fact that both indicators of inauthenticity are present. Therefore, I introduce a clause which delimits the kind of people and desires which are justifiably targeted by judgments of inauthenticity in the present framework. The delimiting clause is that judgments of inauthenticity here only target desire-holders who are known to carry a general wish to live according to the prevailing social and moral standards, and desires that are seriously undesirable according to those standards. The full theory can then be summarized as:

> For persons who wish to live according to the prevailing social and moral standards and desires that are seriously undesirable according to those standards, it is justified to judge that a desire is inauthentic to the extent that it is due to causal factors that are alien to the person and to the extent that it deviates from the person's practical identity.

However, since the publication of the article, I have had to revise my views about the delimiting clause. It may unnecessarily introduce some problems that should be avoided; there is no need to bring the prevailing social and moral standards into the theory. In footnote 4 in the article, I write:

> One plausible line of thought is instead that judgments of inauthenticity may be justified in either case, but that they are only interesting when the desire under scrutiny is bad in some sense.

I now think that this view is better, mainly for reasons of simplicity. Instead of delimiting the scope of desires and desire-holders according to the prevailing social and moral standards, I think that the scope should be delimited to concern only desires which are held by people who may hurt themselves or others. In practice, there may not be a real difference between the two suggestions. The meaning of "hurt," for instance, depends on the prevailing social and moral standards. But, a clause which delimits judgments of inauthenticity to concern only desires held by desire-holders who may hurt themselves or others is more in line with medical practices of compulsory care, which is ultimately the context to which my thesis aims to contribute. Therefore, in the Autonomy article, I build on the revised delimiting clause instead.

Article 3: The Autonomy article

In this article, I add the theory delineated in the Inauthenticity Judgments article to Beauchamp and Childress's theory of autonomy. The result is a non-ideal authenticity-based conceptualization of personal autonomy. I apply it to a paradigm case of possible inauthenticity to test the theory and show that it can provide action-guidance in practical contexts.

Beauchamp and Childress have developed a non-ideal theory of autonomy building on the premise that everyday choices of generally competent persons are autonomous. In the theory, autonomous actions are analyzed "in terms of normal choosers who act (1) intentionally, (2) with understanding, and (3) without controlling influences that determine their action" (2013, p. 104). In the Autonomy article, I add a fourth condition of authenticity to Beauchamp and Childress's theory, building on the arguments in the Inauthenticity Judgments article.

After having expanded Beauchamp and Childress's conceptualization of personal autonomy to include judgments of inauthenticity, I apply the resulting theory to a case of anorexia nervosa. This further develops the theory by demonstrating how it is intended to be applied in practical contexts.

One problem for the application of the theory is that there are no in-depth individual case-descriptions focusing on anorexia nervosa in the literature on authenticity. Therefore, in the article, I construct a hypothetical case from two interview studies which are commonly considered to be authoritative in this context, namely Hope et al. (2011) and Tan et al. (2006). I take citations from real patients and let a hypothetical person, "Amy," represent them. In the article, Amy tells her medical story, which is complex, vague, and contains little detailed information. To the best of my knowledge, it is a realistic description of a person who has been diagnosed with anorexia nervosa.

The analysis of Amy shows that my proposed theory yields reliable results in real cases. Furthermore, it places the notion of authenticity in a conceptual context that is familiar to theorists and practitioners, showing that practical bioethics can encompass ideals of authenticity.

Article 4: The Nine Cases article

In this article, I have collected nine examples of authenticity-related problems in biomedical contexts. Its main merit is that it provides an overview of such problems, and that it points out the limitations of the theory developed in the Autonomy article. Against this background, I argue that there is no universal theory of authenticity which can be applied to solve all authenticity-related problems; the problems require different approaches. Furthermore, I suggest more briefly that authenticity theorists should consider a non-ideal methodological grip on the problems.

The cases collected in the article are both real and hypothetical. Most of them are taken from the bioethical literature on authenticity, but some are taken from conversations with psychiatrists and philosophers. The cases are (1) inauthenticity from physical causes, (2) inauthenticity from psychological causes, (3) unstable desire-sets, (4) lack of desires, (5) medically induced authenticity, (6) inauthentic recovery, (7) indoctrinated desires, (8) false selves, and (9) unexplained surprising desires. Cases 1 through 5 build on actual cases while cases 6 through 9 are hypothetical.

Case 1 describes a 40-year old man who developed pedophilic symptoms that were later found to be causally linked to a brain tumor (Burns and Swerdlow 2003). I use this as a paradigm case of inauthenticity in the Inauthenticity Judgments article. Case 2 treats anorexia nervosa, which has been described already in the above. Case 3 concerns patients suffering from BPD. As mentioned briefly above, BPD patients can sometimes display sudden and dramatic volitional shifts that have been analyzed in terms of authenticity before (Lester 2009). In case 4, I describe persons suffering from late stages of schizophrenia, which may include "negative" symptoms such as passivity and blunting of affect (American Psychiatric Association 2013). Some schizophrenics can be described as living without any desires, and this condition could potentially be analyzed in terms of inauthenticity. Case 5 reproduces a case description from Kramer (1993). The case is a woman who had suffered from severe depression before being successfully treated with Prozac. The woman claims that she is not herself when she is not taking the medicine, which calls for analyses of medically induced authenticity.

The hypothetical cases begin with case 6, which discusses the possibility that there is a difference in terms of authenticity between treating a disorder with medicine and treating it with some other kind of therapy. It is a feasible idea that one recovery process is more authentic than the other. Case 7 is the concern that desires may sometimes be indoctrinated. For instance, a person who grows up in a religious sect and is manipulated into adopting some extreme worldview may have inauthentic desires. Case 8 introduces a thought example which, just as anorexia nervosa, has been considered as a paradigm case of inauthenticity, namely a person who fully conforms to the demands and expectations of others rather than being motivated from his own self (Velleman 2002, p. 97). Finally, case 9 builds on a thought example that I formulate in the Determining Authenticity article, namely Anna, the professional ballet dancer.

I conclude in the article that the problems concern authenticity in different ways and from different perspectives. Some of the problems, namely (1), (2), and (9), should be phrased in terms of authentic decision-making. These can generally be treated with the theory developed in the Autonomy article. Other problems, on the contrary, should be phrased in terms of being an authentic person or being in an authentic condition. These need to be treated with some other theoretical approach than that of this thesis. Therefore, I argue that there is no universal solution to authenticity-related problems but rather various particular solutions, some of which are yet to be treated by applied ethicists.

References

Ahlin, J. (2017). *Personal Autonomy and Informed Consent: Conceptual and Normative Analyses* (licentiate thesis). KTH Royal Institute of Technology, Stockholm, Sweden.

—. (2018a). What Justifies Judgments of Inauthenticity? *HealthCare Ethics Committee Forum* 30(4), 361–377.

—. (2018b). The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent. *Medicine, Health Care and Philosophy* 21(1), 43–50.

Ahlin Marceta, J. (2018). A Non-Ideal Authenticity-Based Conceptualization of Personal Autonomy. *Medicine, Health Care and Philosophy* DOI:doi.org/10.1007/s11019-018-9879-1.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5 ed.). Washington, DC: APA.

Appelbaum, P. S., Lidz, C. W., & Klitzman, R. (2009). Voluntariness of Consent to Research: A Conceptual Model. *Hastings Center Report* 39(1), 30–39.

Beauchamp, T. L., & Childress, J. F. (1983). *Principles of Biomedical Ethics* (2 ed.). Oxford University Press.

—. (2013). *Principles of Biomedical Ethics* (7 ed.). Oxford University Press.

Beauchamp, T. L., & Rauprich, O. (2016). Principlism. In H. ten Have (Ed.), *Encyclopedia of Global Bioethics* (pp. 2282–2293). Springer.

Berghmans, R. (2011). Voluntary Consent, Normativity, and Authenticity. *The American Journal of Bioethics* 11(8), 23–24.

Bialystok, L. (2014). Authenticity and the Limits of Philosophy. *Dialogue* 53(2), 271–298.

Buchanan, A. E., & Brock, D. W. (1990). *Deciding for Others: The Ethics of Surrogate Decision Making*. Cambridge University Press.

Burns, J. M., & Swerdlow, R. H. (2003). Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign. *Archives of Neurology* 60(3), 437–440.

Charland, L. C. (2001). Mental Competence and Value: The Problem of Normativity in the Assessment of Decision-Making Capacity. *Psychiatry, Psychology and Law* 8(2), 135–145.

Christman, J. (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies* 117(1), 143–164.

—. (2009). *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge University Press.

—. (2015). Autonomy in Moral and Political Philosophy. *Stanford Encyclopedia of Philosophy*. Retrieved (December 30, 2018) from http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/

Cudney, P. (2014). What Really Separates Casuistry From Principlism in Biomedical Ethics. *Theoretical Medicine and Bioethics* 35(3), 205–229.

Daniels, N. (2016). Reflective Equilibrium. *Stanford Encyclopedia of Philosophy*. Retrieved (October 10, 2018) from https://plato.stanford.edu/archives/fall2018/entries/reflective-equilibrium/

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Erman, E., & Möller, N. (2013). Three Failed Charges Against Ideal Theory. *Social Theory & Practice* 39(1), 19–44.

Eyal, N. (2012). Informed Consent. *Stanford Encyclopedia of Philosophy*. Retrieved (January 9, 2019) from http://plato.stanford.edu/entries/informed-consent/

Faden, R., & Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. Oxford University Press.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1), 5–20.

Gillon, R. (2003). Ethics Needs Principles—Four can Encompass the Rest—and Respect for Autonomy Should Be "First Among Equals". *Journal of Medical Ethics* 29(5), 307–312.

Grisso, T., Appelbaum, P. S., & Hill-Fotouhi, C. (1997). The MacCAT-T: A Clinical Tool to Assess Patients' Capacities to Make Treatment Decisions. *Psychiatric Services* 48(11), 1415–1419.

Hamlin, A., & Stemplowska, Z. (2012). Theory, Ideal Theory and the Theory of Ideals. *Political Studies Review* 10, 48–62.

Hasan, A., & Fumerton, R. (2016). Foundationalist Theories of Epistemic Justification. *Stanford Encyclopedia of Philosophy*. Retrieved (October 10, 2018) from https://plato.stanford.edu/archives/win2016/entries/justep-foundational/

Hope, P. T., Tan, D. J. O. A., Stewart, D. A., & Fitzpatrick, P. R. (2011). Anorexia Nervosa and the Language of Authenticity. *Hastings Center Report* 41(6), 19–29.

Jonsen, A. R. (2000). *A Short History of Medical Ethics*. Oxford University Press.

Juth, N. (2005). *Genetic Information: Values and Rights. The Morality of Presymptomatic Genetic Testing*. Acta Universitatis Gothoburgensis.

Kramer, P. D. (1993). *Listening to Prozac: A Psychiatrist Explores Antidepressant Drugs and the Remaking of the Self*. Penguin Books.

Lee, M. J. H. (2010). The Problem of 'Thick in Status, Thin in Content' in Beauchamp and Childress' Principlism. *Journal of Medical Ethics* 36(9), 525–528.

Lester, R. J. (2009). Brokering Authenticity Borderline Personality Disorder and the Ethics of Care in an American Eating Disorder Clinic. *Current Anthropology* 50(3), 281–302.

de Maagt, S. (2017). Reflective Equilibrium and Moral Objectivity. *Inquiry* 60(5), 443–465.

Nelson, R. M., Beauchamp, T. L., Miller, V. A., Reynolds, W., Ittenbach, R. F., & Luce, M. F. (2011). The Concept of Voluntary Consent. *The American Journal of Bioethics* 11(8), 6–16.

Noggle, R. (2005). Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (pp. 87–108). Cambridge University Press.

Nozick, R. (1969). Coercion. In W. Morgenbesser (Ed.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel* (pp. 440–472). St Martin's Press.

Oshana, M. A. L. (2015). Is Social-Relational Autonomy a Plausible Ideal? In M. A. L Oshana (Ed.), *Personal Autonomy and Social Oppression: Philosophical Perspectives* (pp. 3–24). Routledge.

O'Neill, O. (1987). Abstraction, Idealization and Ideology in Ethics. In J. D. G. Evans (Ed.), *Moral Philosophy and Contemporary Problems* (pp. 55–69). Cambridge University Press.

—. (2003). Some Limits of Informed Consent. *Journal of Medical Ethics* 29(1), 4–7.

Rauprich, O. (2011). Specification and Other Methods for Determining Morally Relevant Facts. *Journal of Medical Ethics* 37(10), 592–596.

Rawls, J. (2001). *Justice as Fairness: A Restatement* (E. Kelly Ed.). The Belknap Press of Harvard University Press.

Schmidtz, D. (2006). *Elements of Justice*. Cambridge University Press.

Sen, A. (2006). What Do We Want from a Theory of Justice? *The Journal of Philosophy* 103(5), 215–38.

Singer, P. (1993). *Practical Ethics*. Cambridge University Press.

Sjöstrand, M., & Juth, N. (2014). Authenticity and Psychiatric Disorder: Does Autonomy of Personal Preferences Matter? *Medicine, Health Care and Philosophy* 17(1), 115–122.

Strong, C. (2000). Specified Principlism: What is it, and Does it Really Resolve Cases Better than Casuistry? *Journal of Medicine and Philosophy* 25(3), 323–341.

Tan, D. J. O. A., Hope, P. T., Stewart, D. A., & Fitzpatrick, P. R. (2006). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. *Philosophy, Psychiatry, & Psychology: PPP* 13(4), 267–282.

Taylor, J. S. (2005). Introduction. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (pp. 1–29). Cambridge University Press.

—. (2009). *Practical Autonomy and Bioethics*. Routledge.

Valentini, L. (2012). Ideal vs. Non-ideal Theory: A Conceptual Map. *Philosophy Compass* 7(9), 654–664.

Velleman, D. J. (2002). Identification and Identity. In S. Buss & L. Overton (Eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (pp. 91–123). The MIT Press.

Wall, E. (2001). Voluntary Action. *Philosophia* 28(1–4), 127–136.

Wiens, D. (2015). Against Ideal Guidance. *Journal of Politics* 77(2), 433–46.

van Willigenburg, T. (2007). Reflective Equilibrium as a Method in Health Care Ethics. In R. E. Ashcroft, A. Dawson, H. Draper, & J. R. McMillan (Eds.), *Principles of Health Care Ethics* (pp. 207–213). John Wiley & Sons.

Wolff, J. (2011). *Ethics and Public Policy: A Philosophical Inquiry*. Routledge.

Part II

Articles

The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent

Abstract: It is sometimes argued that autonomous decision-making requires that the decision-maker's desires are authentic, i.e., "genuine," "truly her own," "not out of character," or similar. In this article, it is argued that a method to reliably determine the authenticity (or inauthenticity) of a desire cannot be developed. A taxonomy of characteristics displayed by different theories of authenticity is introduced and applied to evaluate such theories categorically, in contrast to the prior approach of treating them individually. The conclusion is drawn that, in practice, the authenticity of desires cannot be reliably determined. It is suggested that authenticity should therefore not be employed in informed consent practices in healthcare.

Keywords: Authenticity, autonomy, informed consent, decision-making, healthcare

Introduction

Informed consent is a patient's valid authorization or refusal of a medical intervention; a process aiming at protecting patients' autonomy. In its elaborate form it is usually understood as informed, voluntary, and competent consent (cf. Eyal 2012). Clinicians sometimes meet patients who are competent, yet make (at least seemingly) incomprehensible treatment decisions.[1] Some of those decisions can be described as *inauthentic*.

The question can be raised whether the authenticity of decisions should be included as a criterion in informed consent to further pro-

[1] Competent according to, e.g., the MacCAT-T: a clinical tool to assess patients' capacities to make treatment decisions. See Grisso et al. (1997).

tect patients with regards to their autonomy.[2] In this article, I argue that the authenticity (or inauthenticity) of desires cannot be reliably detected. Therefore, authenticity should not be part of informed consent. A well-founded suspicion that a desire is inauthentic may call for other measures than the invalidation of consent (or refusal), such as a moral obligation to double-check that the patient is competent to make healthcare decisions. However, the aim of this article is mainly theoretical. Although some possible policy implications are suggested, none is defended at length.

The paper is structured as follows. In "The problem of authenticity and informed consent," I elaborate on the problem of authenticity and informed consent. In "A taxonomy of authenticity," I introduce a taxonomy of characteristics displayed by theories of authenticity. In "The taxonomy and the argument from testability," I use the taxonomy to evaluate the prospect of different theories of authenticity to produce reliably observable consequences. Lastly, "Concluding remarks" contains some concluding remarks.

The problem of authenticity and informed consent

Anna

Consider the hypothetical case of Anna, a young and promising professional ballet dancer. Anna loves her work. She has moved across the nation to attend the best ballet schools, set aside personal relationships when they conflict with her career, and is known by friends and family to love dancing "more than anything else." Anna has suffered a serious leg injury. To avoid the risk of having to go through an amputation that will definitely end her career as a dancer, she must undergo a minor surgery. She understands information relevant to her condition, is capable to reason about the potential risks and benefits of different treatment

[2] See, e.g., O'Shea, who raises the possibility of introducing authenticity as a necessary condition of consent in order to distinguish between benign persuasion and undue influence (2011, pp. 30–1).

alternatives, appreciates the nature of her situation, the consequences of her choices, and so on. Yet, she refuses to undergo surgery.

There is no physiological or psychological disorder, such as a brain tumor, untreated syphilis, or psychosis, that can be tied to Anna's decision-making. Neither is she being forced or unduly influenced to make a decision that accords with someone else's interests, certain social relations, authoritative traditions, or anything else that might impinge on the voluntariness of her choices. She plainly refuses to undergo surgery.

When reflecting upon the case, her doctor considers Anna's treatment decisions to be "out of character." She believes that Anna is not being "herself," which is why she makes choices that are not "genuine." Nonetheless, the doctor must conclude, Anna is competent to make treatment decisions; nothing in the informed consent process would allow anyone to override Anna's choices. However, if informed consent had included a criterion of *authenticity*, Anna's decisions could have been invalidated on that basis. Her "true wishes" could then be adhered to by forcing the measures necessary to save Anna from amputation. Therefore, the doctor contemplates whether or not the authenticity of patients' decisions should be part of informed consent.

The question arises in various contexts. For instance, anorexia nervosa patients have stated that the disorder "was a part of themselves, and therefore it would not be them if they recovered" (Tan et al. 2006b, p. 278). Similarly, some people with bipolar disorder have been reported to ask whether certain experiences are due to their illness, medication, or themselves (Hope et al. 2011, p. 21). And, brain tumors can entail personality changes, such as in the case of a 40-year-old male who suddenly developed pedophilia (Burns and Swerdlow 2003). All examples of cases in which the concept of authenticity can be invoked.

Authentic desires and informed consent

There are several interrelated problems concerning the question of whether the authenticity of patients' decisions should be part of informed consent. First, it must be determined what authenticity is. Lexical definitions of "authentic" include descriptions such as "real or gen-

uine," "not copied or false," "true and accurate," and so on, but for moral reasons it is necessary to adopt a more detailed and systematized account, i.e., a *theory* of authenticity.[3] Second, a method must be developed that enables observers to reliably recognize authenticity (or inauthenticity) in others. Merely having a theory of authenticity does not suffice, as the concept is (or is not) to be applied in a context in which interpersonal morality requires that interventions with other people's lives and liberties are justified. It is first when these two matters are satisfactorily settled that we are in a position to judge whether or not to include authenticity in informed consent.

This article treats the second of the above stated problems. Thus, I do not aim to contribute to the philosophy of authenticity—although I believe that this work does so indirectly—but merely to its applicability. I claim to show that a method that enables observers to reliably recognize authenticity (or inauthenticity) in others cannot be developed. However, this claim must be conditioned. First, I only take into consideration theories of authenticity present in contemporary literature on personal autonomy. Second, my claim is delimited by the fact that I only treat theories in what is commonly called a *procedural* tradition of personal autonomy, which can be contrasted with a *substantial* tradition. In the procedural tradition, theorists are only concerned with the process by which desires are formed and realized. In the substantial tradition, theorists are also concerned with the content of a desire-holder's desires (see e.g. Oshana 2015). Third, I assume that authentic desires can be treated without a well-articulated idea of what it is to be an authentic *person*. This assumption requires some elaboration.

Much of what has been said of authenticity is phrased as "preferences stemming from her *true self*," and similar. The problem with such phrases is that they necessitate some idea of personhood. In the humanities, it is a frequently debated problem what personhood is, or what it is to be a person. Are we socially constituted beings, as some believe, or are we

[3] These descriptions are from Merriam-Webster online. The arguments in this article do not commit to any specific lexical definition of "authenticity," but treats a number of suggestions that have been proposed with regards to how the concept should be understood.

self-made? Is *tabula rasa* a real thing? And, in all cases, to what extent? I think that the current problem is possible to treat without engaging in such debates. That is, it should not matter to my argument or to informed consent whether humans are socially constructed beings or if we are something else. Whatever we are, I am here concerned only with *desires*. In this context, I intend for desires to be understood as the basic element in preference forming, i.e., basic pro-attitudes. Therefore, I treat theories of authenticity as theories of authentic desires—although these often include a mix of propositions about "authentic selves," "authentic decisions," "authentic preferences," and so on.

Method

I approach the problem as follows. Sjöstrand and Juth recently concluded that the concept of authenticity is "highly problematic to use as a criterion for autonomous decision-making in healthcare" (2014, p. 115). Although I agree with them, it is not my intention to merely reproduce their arguments here. I wish to strengthen their conclusion with new arguments. Sjöstrand and Juth only treat authenticity in the context of psychiatric care. However, I use a method that allows me to conclude that authenticity is problematic in the above sense in all healthcare settings. My method requires a more in-depth explanation of the problem at hand.

Sjöstrand and Juth write the following (p. 121):

> The practical question is which patients should be deemed inauthentic enough not to be granted certain rights typically granted to patients considered fully autonomous—for instance, a right to refuse treatment. Hence, we also need to have some idea about how to *test* patients regarding the authenticity of their desires. This seems to be very difficult…

I call this *the argument from testability*. It is further developed in "The argument from testability." Here, it suffices to declare that it is more significant than Sjöstrand and Juth acknowledges. First, testability is central to the problem of developing a method that enables observers to

reliably recognize authenticity (or inauthenticity) in others. Second, the argument from testability applies in some form not only to the theory of authenticity favored by Sjöstrand and Juth. If my thesis holds, the argument from testability applies universally, and authenticity cannot be reliably employed as a criterion in informed consent practices.

As stated above, I use a different method than Sjöstrand and Juth's. They go through a collection of theories of authenticity individually and demonstrate in each case how that specific theory is flawed. One problem with that method is that it is space consuming. It requires of the authors to briefly summarize each theory—which paves the way for misrepresentations—and to, just as briefly, demonstrate precisely what is wrong with it. Another problem is that many theories may be left out of the analysis. By contrast, in this article, I introduce a taxonomy of characteristics displayed by different theories of authenticity that allows me to treat such theories categorically. The method is less space consuming and its results more reliable, although it cannot be guaranteed that the taxonomy covers all conceivable characteristics of authenticity. Nonetheless, my method collects many theories of authenticity, several of which have been highly influential, and makes their similarities and differences comprehensible.[4] Even if my conclusion is unconvincing, the taxonomy is still a valuable contribution to the discussion of authenticity in autonomy theory.

A taxonomy of authenticity

The taxonomy

There are many theories of authenticity.[5] As is made clear above, I will not attempt to go through them all here. However, I will account for some distinctive elements that many theories share. This allows me to organize characteristics displayed by different theories of authenticity

[4] I am not aware of any theory that the taxonomy does not cover.

[5] In addition to those explicitly mentioned in this article, see, e.g., Buchanan and Brock (1990), Chariand (2001), DeGrazia (2005), Faden and Beauchamp (1986), Freedman (1981), Tännsjö (1999), Velleman (2002), Winnicott (2007).

into tree distinct categories: *sanctionist*, *originist*, and *coherentist*. These are not formal definitions, but broad categories that distinguish different conceptualizations of authenticity. In sanctionist theories, i.e., theories distinguished by characteristics typical of sanctionist ideals, authenticity concerns the desire-holder's attitude towards her desires. In originist theories, authenticity concerns the origin of a desire. In coherentist theories, authenticity concerns the coherence of a desire-holder's set of desires. This will be elaborated below. Furthermore, these categories come in two classes: *cognitivist* and *non-cognitivist*. In cognitivist theories, authenticity is a matter of rational deliberation; non-cognitivist theories do not commit to that. Thereby, non-cognitivist theories do not reject rational deliberation altogether, they merely do not commit to the narrow view that authenticity is only a matter of rational deliberation. A theory of authenticity can display characteristics from more than one category. The classes on the other hand are mutually exclusive, so that a theory is either one or the other.

Thereby, the taxonomy takes the form of a three-by-two scheme.[6] I will go through the different categories and classes respectively, and illustrate their distinct features by using quotes and examples from theories that display elements that are characteristic for each category and class.

Sanctionism

In sanctionist theories, authenticity concerns the desire-holder's attitude towards her desires. Desires that in one way or another are sanctioned by the desire-holder are deemed authentic. Consider, for instance, Frankfurt, whose idea of a *person* is that such a being identifies reflectively with her desires, and Dworkin, who holds that it is crucial to a person's autonomy that she has the "capacity to raise the question of whether [she] will identify with or reject the reasons for which [she] now act[s]" (Frankfurt 1971, pp. 10–17; Dworkin 1988, p. 15). Similarly, Juth writes that "the most important property of an authentic desire is that a person who has the desire would

[6] A third dimension could be added to the taxonomy, marking the degree to which a theory displays the characteristic in question. However, my argument does not require such elaborations and it will therefore be left out of the analysis.

be inclined to approve of having that desire if she came to know why she has it" (2005, p. 129). This is also the type of theory that Sjöstrand and Juth favors: it is "the person's own attitude towards the desire in the light of knowledge about the origin that matters" (2014, p. 121).

According to sanctionist theories, the status of a desire in terms of authenticity is determined by means of endorsement. Suppose that Anna came to know exactly why she has the desire to refuse to undergo the minor surgery that is necessary to avoid the risk of amputation. In this hypothetical state of mind, she is aware of everything that might subconsciously influence her desire forming; nothing regarding her psychological and physiological behavioral patterns escapes her internal gaze. Sanctionist theories suggest that Anna's desires are authentic if and only if Anna, in this hypothetical state of mind, would endorse the reasons for why she has the desire in question.

The above are examples of *cognitivist* sanctionist theories of authenticity. According to them, authenticity is a matter of rational deliberation. Frankfurt suggests that persons identify *reflectively* with their desires and Dworkin writes about a "capacity to *raise the question*" (emphasis added; see quote above). Accordingly, Sjöstrand and Juth use the label "*Rationally* endorsed desires" to describe theories such as these (p. 120; emphasis added). I know of no *non-cognitivist sanctionist* theories, but the taxonomy may allow us to formulate one. A theory could, perhaps, be developed so that a desire is authentic only if the desire-holder experiences an emotional inclination in favor of it.

Originism

In originist theories, authenticity concerns the origin of a desire. In a manuscript, Tan et al. formulate an originist theory of authenticity as a counterfactual statement: Authentic views are such that a person "would have (or did have) if she did not suffer from [a disorder]" (2006a, p. 20).[7] Similarly, but more elaborately, Elster argues when writing about the rationality of desires that desires are inauthentic if they are "shaped by

[7] This is omitted in the published version of the article (Tan et al. 2006b).

irrelevant causal factors, by a blind psychic causality operating 'behind the back' of the person" (1983, p. 16; Sjöstrand and Juth 2014, p. 118). All desires have a "causal origin, but some of them have the wrong sort of causal history" (Elster 1983, p. 16). Elster continues by writing about persons that "are in control over the processes whereby their desires are formed," stating that "autonomous [here: authentic] desires are desires that have been deliberately chosen, acquired or modified—either by an act of will or by a process of character planning" (p. 21). Thus, according to Elster, authentic desires are such that originate in some cognitive process controlled by the desire-holder. That is, Anna's desire to refuse to undergo surgery to avoid the risk of amputation could originate in something that is beyond her cognitive control.

An example of an originist theory of authenticity that can be interpreted as non-cognitivist is found in Meyers. Arguing against Frankfurt (see above), Meyers writes that having "an authentic self is best understood as the result of an ongoing activity of persons" (2001, p. 199). The authentic self is "the evolving collocation of attributes—analogous to a musical ensemble's sound—that issues from ongoing exercise of" a repertory of skills of "introspection, imagination, memory, communication, reasoning, interpretation, and volition" that enable self-discovery and self-definition (ibid). Elsewhere, Meyers writes that when exercising such skills one "*enacts* one's authentic self" (2005, p. 49). Although the theory is built on a cognitivist foundation, it is ultimately non-cognitivist. Meyers writes that what "autonomous people do to understand and define themselves is not aptly figured by any Euclidean shape or formal reasoning pattern" (2001, p. 199). Thus, enacting one's authentic self is not a rationalist enterprise. A Meyerean theory of authenticity phrased in terms of desires could be formulated accordingly: desires are authentic if and only if they originate in non-cognitivist processes of self-discovery and self-definition.

Coherentism

In coherentist theories, authenticity concerns the coherence of a desire-holder's set of desires. Christman argues that for a characteristic to be authentic it must pass a self-critical reflection, similar to that in cognitivist sanctionist theories. However, the reflection does here not target the rational endorsement of having a certain desire, but whether the characteristic in question can be "sustained as part of an acceptable autobiographical narrative organized by her diachronic practical identity" (2009, p. 155). While sanctionism is an atomist theory focusing on individual desires, coherentism is holist; authenticity here concerns a whole body of desires.

Phrased in terms of desires, a Christmanean theory of authenticity could be that a person's desires are authentic if and only if they fit with her socio-historical or autobiographical narrative. Anna's desire to refuse to undergo surgery does not fit with her socio-historical or autobiographical narrative. She loves do dance "more than anything else," is known to have set aside personal relationships when they have conflicted with her career, and so on. Her present desires just do not *fit*.

The Christmanean theory is cognitivist. Similarly, albeit as an example of a non-cognitivist coherentist theory, Miller writes (1981, p. 24):

> Autonomy as authenticity means that an action is consistent with the person's attitudes, values, dispositions, and life plans. Roughly, the person is acting in character. [...] For an action to be labeled "inauthentic" it has to be unusual or unexpected, relatively important in itself or its consequences, and have no apparent or proffered explanation.

These are the categories and classes of characteristics displayed by different theories of authenticity. Below, the taxonomy is used to test such theories categorically.

The taxonomy and the argument from testability

The argument from testability

Most propositions and theories can be tested in several ways. One test could, for instance, aim at identifying conceptual vaguenesses, ambiguities, and inconsistencies in theories of authenticity. The concern of the argument from testability, however, is something else. Theories of authenticity will here not be evaluated as such. Since authenticity is (or is not) to be applied in informed consent contexts, it is a necessary criterion of a theory of authenticity that it renders observable and testable consequences. Therefore, it is only the prospect of the theories producing empirically observable consequences, and the possibility of evaluating those consequences, that is of interest here. Contemporary theories of authenticity may be good in other respects, although it is beyond the present purpose to assess that.

The taxonomy of characteristics displayed by different theories of authenticity allows us to evaluate the testability of theories of authenticity categorically. If it is true that neither sanctionist, originist, nor coherentist characteristics can produce observable and testable consequences, no theory that builds on those elements and those elements only achieves the requirement posed by the argument from testability. In "Sanctionism" through "Coherentism," I spell out what the argument from testability requires of each category of characteristics, and show that no such category passes the test.

Sanctionism

Suppose that Anna's doctor is a sanctionist regarding authenticity. She believes that for a desire to be authentic it must be hypothetically endorsed by the desire-holder. There are two main reasons why this view does not render any observable and testable consequences. First, as Sjöstrand and Juth write (p. 121):

> For one thing, it is often difficult to come up with a full explanation as to why we have a certain desire, and even more difficult to

> make the necessary investigations in order to determine whether or not this explanation is correct.

This practical problem may be overcome, as discussed in "Originism" below. But, in sanctionist theories, desire-holders are to transcend into a state of mind from which the status of a desire is assessed. There are two possibilities here. Either that state of mind is hypothetical, in which case the theory cannot render observable consequences (but merely hypothetical ones). Sanctionist theories are then not falsifiable. Or, it is an actual state of mind. If it is an actual state of mind, observers must evaluate whether the desire-holder transcends into *it*, into some *other* state of mind, or if she does not transcend into anything at all. Furthermore, they must reliably determine whether valid endorsement is actually taking place when the desire-holder is in that state of mind. To do so would require access to advanced (and currently unavailable) neuro-imaging technology, in addition to an in-depth knowledge of the psychological nature of endorsement. It would appear that sanctionism is, at the very least, impractical.

In conclusion, sanctionism does not render observable and testable consequences without technology and scientific knowledge yet unheard of, if at all. That entails that, at least as of today, a method that enables observers to reliably recognize authenticity (or inauthenticity) in others cannot be grounded in sanctionist theories of authenticity only.

Originism

Suppose instead that Anna's doctor is an originist regarding authenticity. She believes that for a desire to be authentic it must originate in a process controlled by the desire-holder. In practice, this view also fails to render observable and testable consequences.[8]

Again quoting Sjöstrand and Juth, it is difficult "to come up with a full explanation as to why we have a certain desire, and even more difficult to

[8] As pointed out by an anonymous reviewer, the question of tracing the origins of a desire may, at least partially, be metaphysical rather than (socio-)psychological. That may be true, but metaphysical theses are not empirically testable, so I choose here not to address the possibility of metaphysical origins of desires.

make the necessary investigations in order to determine whether or not this explanation is correct." Observers face the insurmountable task of tracing the origin of desires in hindsight and attempt to reliably determine when they were formed. And, if that problem is resolved and the time of origin detected, observers must also reliably determine whether the desire-holder was in control over the desire-forming process at the time.

These problems are significant in theory, but plausibly impossible to overcome in practice. Against scarcity of resources, healthcare practices would have to develop manageable and effective methods to examine the origin of desires. Among other things, those methods would likely have to include deep psychological analysis and a detailed socio-historical biographical investigation. In addition to that, to determine whether the desire-holder was in control of the desire-forming process, it is likely the case that the methods would have to include interviews with people who were close to the desire-holder when the desires were initially formed, and other similar measures. To complicate things further, these investigations would also require the desire-holder's informed consent.

To conclude, originist theories may render observable and testable consequences in theory. However, to examine the matter would require overwhelmingly complex and resource-demanding methods. Therefore, it is plausibly insurmountably difficult for healthcare practices to reliably recognize originist authenticity (or inauthenticity) in patients.

Coherentism

Suppose, then, that Anna's doctor is a coherentist regarding authenticity. She believes that authenticity concerns the coherence of a desire-holder's set of desires. Naturally, she thinks of Anna's desire to refuse to undergo minor surgery to avoid the risk of amputation as diverging. In short, the desire does not fit.

Assessing the authenticity of Anna's desire requires an exhaustive list of her desires. In addition to her desire to move across the country, attend the best ballet schools, and set aside personal relationships that conflict with her career, it must include desires that may arise in situations not immediately or obviously connected to the present one.

The set must also include desires in unknown situations, e.g., such that will arise in the future and of which nothing can be known. It cannot be determined when a desire-set is full. Therefore, observers cannot reliably determine the coherence of a specific desire.

Prima facie, a reasonable way to circumvent the problem of composing an exhaustive desire-set is to in some way delimit the extent of the set, although a reflected judgment reveals that doing so implies making normatively substantial decisions. Delimiting the set necessitates deciding that some desires are irrelevant to the assessment. In fact, coherentism is inherently normative (cf. Banner and Szmukler 2013, p. 390). It cannot be explained why a diverging desire is inauthentic rather than the rest of the desire-holder's set of desires, without invoking the support of normative auxiliary assumptions. That is, Anna's doctor cannot be sure that it is not Anna's desires to move across the country, attend the best ballet schools, and set aside personal relationships that conflict with her career that are inauthentic. Empirical data, or incoherency as such, do not reveal which piece of the desire pie that should be discarded; the large or the small one. The truth of the matter cannot be *discovered*, but must be *decided*.

An intuitively compelling example that corresponds to the case of Anna is a person who suddenly reveals that she is homosexual, to the surprise of everyone close to her. Her romantic desire toward others of the same sex cannot be thought of as "inauthentic" only because it deviates from her previously displayed desires, unless some normative auxiliary assumption is invoked in favor of the largest piece of the desire pie. Therefore, coherentism is an inherently normative characteristic in authenticity theory.

In conclusion, even if the problem of composing an exhaustive desire-set is overcome, coherentist characteristics do not render observable and testable consequences independent from normative auxiliary assumptions. Therefore, a method that enables observers to reliably recognize authenticity (or inauthenticity) in others cannot be grounded in coherentist theories of authenticity only; it also requires a moral defense.

Concluding remarks

Above, it has been shown that theories that build on characteristics covered by the taxonomy fail to meet the requirements set by the argument from testability. However, that does not imply that we can be sure that authenticity cannot be part of informed consent. There might be characteristics and theories that the taxonomy here introduced does not cover. Furthermore, my assumption that authentic desires can be analyzed without a well-articulated idea of authentic persons may be mistaken. The same applies to my choice to only treat theories of authenticity in the procedural tradition of personal autonomy theory. Substantial theories of authenticity have been left out of the present analysis; they may succeed where procedural theories do not. Lastly, the alternative remains to begin with what can be reliably detected regarding desires and develop a theory of authenticity thereafter—that is, to intentionally put the cart before the horse.

However, if my assumptions are sound and the taxonomy is exhaustive, in practice, the authenticity (or inauthenticity) of desires cannot be reliably detected. Therefore, authenticity should not be included as a criterion in informed consent.

Nonetheless, seemingly inauthentic behavior from patients may trigger the need to take other actions than invalidating consent (or refusal). Anna's doctor may, for instance, be morally obliged to double-check that Anna is able to comprehend the nature of her situation. Or, surprising desires such as Anna's might prompt the need for alternative communicative measures, e.g., the use of pedagogical tools, or perhaps another doctor's affirmation that the information the patient has received is correct. However, it is beyond the limits of this article to further treat moral obligations that may arise from a suspicion of inauthenticity. Any detailed policy suggestions based on the conclusions drawn in this article must be carefully but separately formulated.

References

Banner, N. F., & Szmukler, G. (2013). 'Radical Interpretation' and the Assessment of Decision-Making Capacity. *Journal of Applied Philosophy* 30(4), 379–394.

Buchanan, A. E., & Brock, D. W. (1990). *Deciding for Others: The Ethics of Surrogate Decision Making*. Cambridge University Press.

Burns, J. M., & Swerdlow, R. H. (2003). Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign. *Archives of Neurology* 60(3), 437–440.

Charland, L. C. (2001). Mental Competence and Value: The Problem of Normativity in the Assessment of Decision-Making Capacity. *Psychiatry, Psychology and Law* 8(2), 135–145.

Christman, J. (2009). *The Politics of Persons: Individual Autonomy and Socio-historical Selves*. Cambridge University Press.

DeGrazia, D. (2005). *Human Identity and Bioethics*. Cambridge University Press.

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.

Eyal, N. (2012). Informed Consent. *Stanford Encyclopedia of Philosophy*. Retrieved (October 10, 2018) from http://plato.stanford.edu/entries/informed-consent/

Faden, R., & Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. Oxford University Press.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1), 5–20.

Freedman, B. (1981). Competence, Marginal and Otherwise: Concepts and Ethics. *International Journal of Law and Psychiatry* 4(1–2), 53–72.

Grisso, T., Appelbaum, P. S., & Hill-Fotouhi, C. (1997). The MacCAT-T: A Clinical Tool to Assess Patients' Capacities to Make Treatment Decisions. *Psychiatric Services* 48(11), 1415–1419.

Hope, P. T., Tan, D. J. O. A., Stewart, D. A., & Fitzpatrick, P. R. (2011). Anorexia Nervosa and the Language of Authenticity. *Hastings Center Report* 41(6), 19–29.

Juth, N. (2005). *Genetic Information: Values and Rights. The Morality of Presymptomatic Genetic Testing*. Acta Universitatis Gothoburgensis.

Meyers, D. T. (2001). Authenticity for Real People. In B. Elevitch (Ed.), *Proceedings of the Twentieth World Congress of Philosophy: Philosophy of Mind and Philosophy of Psychology* (pp. 195–202). Philosophy Documentation Center.

—. (2005). Decentralizing Autonomy: Five Faces of Selfhood. In J. Christman & J. Anderson (Eds.), *Autonomy and the Challenges to Liberalism* (pp. 27–55). Cambridge University Press.

Miller, B. L. (1981). Autonomy and the Refusal of Lifesaving Treatment. *The Hastings Center Report* 11(4), 22–28.

O'Shea, T. (2011). Consent in History, Theory and Practice. *Essex Autonomy Project Green Paper Report*. University of Essex: Essex Autonomy Project.

Oshana, M. A. L. (2015). Is Social-Relational Autonomy a Plausible Ideal? In M. A. L Oshana (Ed.), *Personal Autonomy and Social Oppression: Philosophical Perspectives* (pp. 3–24). Routledge.

Sjöstrand, M., & Juth, N. (2014). Authenticity and Psychiatric Disorder: Does Autonomy of Personal Preferences Matter? *Medicine, Health Care and Philosophy* 17(1), 115–122.

Tan, D. J. O. A., Hope, P. T., Stewart, D. A., & Fitzpatrick, P. R. (2006a). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. Retrieved (October 10, 2018) from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2121578/pdf/nihms5623.pdf

—. (2006b). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. *Philosophy, Psychiatry, & Psychology: PPP* 13(4), 267–282.

Tännsjö, T. (1999). *Coercive Care: The Ethics of Choice in Health and Medicine*. Harvard University Press.

Velleman, D. J. (2002). Identification and Identity. In S. Buss & L. Overton (Eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (pp. 91–123). The MIT Press.

Winnicott, D. W. (2007). Ego Distortion in Terms of True and False Self The Maturational Process and the Facilitating Environment. In D. W. Winnicott (Ed.), *Studies in the Theory of Emotional Development* (pp. 140–154). Karnac Books.

 The article is originally published as: Ahlin, J. (2018). The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent. *Medicine, Health Care and Philosophy* 21(1), 43–50. **The following change has been made in the present version:** The list of references has been updated according to the reference standards of this thesis. **Acknowledgements:** I am grateful to three anonymous reviewers, the Higher Seminar at the Division of Philosophy, KTH Royal Institute of Technology, and the *Addressing Ethical Obstacles to Person Centred Care* research program workshop (Gothenburg, October 12–13 2016) for useful comments on earlier drafts of this article.

What Justifies Judgments of Inauthenticity?

Abstract: The notion of authenticity, i.e., being "genuine," "real," or "true to oneself," is sometimes held as critical to a person's autonomy, so that inauthenticity prevents the person from making autonomous decisions or leading an autonomous life. It has been pointed out that authenticity is difficult to observe in others. Therefore, judgments of inauthenticity have been found inadequate to underpin paternalistic interventions, among other things. This article delineates what justifies judgments of inauthenticity. It is argued that for persons who wish to live according to the prevailing social and moral standards and desires that are seriously undesirable according to those standards, it is justified to judge that a desire is inauthentic to the extent that it is due to causal factors that are alien to the person and to the extent that it deviates from the person's practical identity. The article contributes to a tradition of thinking about authenticity which is known mainly from Frankfurt and Dworkin, and bridges the gap between theoretical ideals of authenticity and real authenticity-related problems in practical biomedical settings.

Keywords: Authenticity, autonomy, decision-making, paternalism, bioethics

Introduction

Personal autonomy, i.e., self-determination, is a central notion in contemporary bioethics. Generally speaking, a person is autonomous if she is self-governed. Factors that undermine autonomy include, for instance, lacking decision-making capacities and controlling influences such as coercion or manipulation. Sometimes authenticity, i.e., being "genuine," "real," "true to oneself," or similar, is held as critical to a person's autonomy, so that inauthenticity prevents the person from making autonomous decisions or leading an autonomous life. It is a bioethical

problem how authenticity should be understood. Various theories have been proposed with the intention of conceptualizing authenticity; none takes complete precedence over others.

In practice, the notion is relevant mainly in considerations of what justifies judgments of inauthenticity. For instance, patients suffering from borderline personality disorder (BPD) sometimes display sudden and dramatic shifts in goals, values, vocational aspirations, choice of friends, and so on (Lester 2009, p. 284). During a short time span, a BPD patient can both request medication, as only that enables her to go through psychotherapy, and refuse medication, as one of its side effects is that it clouds her thinking. Healthcare personnel cannot adhere to both wishes. Caretakers with autonomy-promoting or paternalistic ambitions may be interested in whether it is justified to treat any of the BPD patient's decisions as inauthentic, and if so, on what grounds. Other examples include late stage schizophrenics who are completely indifferent to how their lives go (cf. American Psychiatric Association 2013) and anorectics who report that they would rather die than gain weight (Tan et al. 2006). In such cases, it is sometimes relevant to ask whether it is justified to treat patients' wishes as inauthentic.

Thus, there is a three-step problem regarding the notion of authenticity. First, it is unclear under which conditions something or someone is authentic or inauthentic. Second, it is difficult to know whether something or someone meets those conditions. Mainly for reasons of epistemic uncertainty, it is therefore unclear what justifies the judgment that something or someone is inauthentic. Third, some paternalistic or authenticity-promoting interventions may also be justified in light of judgments of inauthenticity. This article is only concerned with the second step of this three-step problem.

I argue that for persons who wish to live according to the prevailing social and moral standards and desires that are seriously undesirable according to those standards, it is justified to judge that a desire is inauthentic to the extent that it is due to causal factors that are alien to the person and to the extent that it deviates from the person's practical identity. My arguments in this article contribute to a tradition of thinking

about authenticity which is mainly known from Frankfurt (1971) and Dworkin (1988) and has recently been supported by Juth (2005) and DeGrazia (2005), among others. However, my contribution is more practical than the theoretical ideals proposed by those authors; this article is an attempt to bridge the gap between theoretical ideals of authenticity and real authenticity-related problems in practical biomedical settings.

The article has two main sections and is structured as follows. In the first main section, I elaborate on how the notion of authenticity is relevant to biomedicine and introduce two recent attempts to collect theories of authenticity in taxonomies. I spell out my proposal of what justifies judgments of inauthenticity in the second main section. My proposal builds on the arguments in the first section, which is why the first section is rather detailed and takes up much space. A brief final section concludes.

Theorizing about authenticity

Authenticity and biomedicine

Autonomy is one of the main guiding principles in contemporary bioethics. In the standard model of autonomy, a person is autonomous with respect to her desires or actions to the extent that they are due to her own self, and not due to some other influencing force, be it internal or external to her (cf. Taylor 2005a). Bioethicists usually invoke two main notions with regard to patients' autonomy; decision-making capacity and voluntariness. That is, if a patient is competent to make healthcare decisions, according to, e.g., MacCAT-T standards of decisional-capacity (Grisso et al. 1997), and does so without undue influences internal or external to the patient (Nelson et al. 2011), most bioethicists agree that the patient's decisions should be respected. Sometimes a third notion is raised, namely that of authenticity. The reasons vary. A patient may make healthcare decisions that seem to be "out of character," or that seem to conform to others' wishes rather than to her own, or she may suffer from some medical disorder that seems to affect her values. Accordingly,

the notion of authenticity has increasingly gained theorists' attention (see, e.g., Bauer 2017; Sjöstrand and Juth 2014; White 2018).

It has become clear that there is no consensus regarding how authenticity should be understood, or what exactly it may add to the concern for patients' autonomy. To the contrary, various theoretical approaches are present in bioethical conversations, none of which takes precedence over others and, arguably, none of which manages to solve all authenticity-related problems that bioethicists have raised. Below, I account for two attempts to collect theories of authenticity in taxonomies that enable overview and analysis. These accounts will be relevant to the arguments in the subsequent section. However, before proceeding, some delimitations are necessary.

The notion of authenticity has been understood to apply to different things. Some argue that it is the authenticity of persons that is of importance to autonomy theory (cf. Bauer 2017). Others hold that it is the authenticity of a person's life that should be considered (cf. Taylor 1991). Although both perspectives are important, I am here concerned with a third possibility, namely, the authenticity of desires. For the present purposes I take desires to be the most basic element in ordinary preference-forming and, thus, a basic element in decision-making. In brief, I hold that autonomy in medical settings mainly concerns decision-making, in the sense that bioethicists are interested in whether patients make autonomous healthcare decisions. Therefore, I phrase concerns of authenticity in terms of the authenticity of decisions, or more precisely in terms of the authenticity of desires.[1]

The notion of authenticity is relevant in several ways. A general theory of authenticity can be applied in common autonomy-protecting practices, such as, e.g., informed consent (cf. Eyal 2012). It is possible that such practices can be developed in light of insights from authenticity theory so that they better protect patient autonomy. But the notion is also relevant for paternalistic reasons. Sometimes, the principle of respect for autonomy is overridden by concerns for a patient's well-being. Although

[1] Also, the focus on desires is common in the theoretical tradition which I aim to contribute to; see, e.g., Noggle (2005), Sjöstrand and Juth (2014), and Taylor (2005b).

compulsory care is rare, it is occasionally considered necessary. And, in some of those cases, the decision to put a patient in compulsory care is made with support from judgments of inauthenticity. That is, patients are sometimes subjected to compulsory care because they display what seems to be inauthentic desires (cf. Tan et al. 2006).

It must be noted that paternalistic interventions are not justified simply because a desire is found to be inauthentic. Paternalism requires support from independent moral arguments, such as the necessary degree of epistemic certainty of inauthenticity and the reasonable proportionality of the intervention. This article does not seek to provide practical guidance in those matters. It is beyond the scope of the present purposes to elaborate more precisely on the relationship between authenticity, autonomy, and paternalism. Here, the only concern is to determine what justifies judgments of inauthenticity.

Finally, I do not claim that my proposal is the only way to justify judgments of inauthenticity. The conclusion is not phrased in terms of necessary and sufficient conditions but should be understood as generally reason-giving in a larger framework of reflective equilibrium. Thus, applying it in practice requires substantial moral deliberation (cf. Beauchamp and Rauprich 2016).

Two taxonomies of authenticity theories

There have been two recent attempts at collecting theories of authenticity in taxonomies; Noggle (2005) and Ahlin (2018). Noggle's taxonomy is important to the arguments in the next section as it distinguishes between so-called procedural and substantive theories of authenticity. The taxonomy I have proposed is important because it fleshes out two different kinds of theories that are conflated in Noggle's taxonomy. As will be explained, these theories are fundamental to the arguments in the next section.

Noggle's taxonomy builds on the observation that theories of authenticity begin with a base clause (2005, p. 88):

> Element (or set of elements) E1 of the psychology of person S is authentic if…

Then, he orders different theories after which conditions they add to complete the clause. Three families of theories emerge (Noggle 2005, p. 88):

> *Structural Condition Schema*: E_1 is related in the right way to E_2, where E2 is some other element (or group of elements) of S's psychology.
> *Historical Condition Schema*: E_1 arose in the right way.
> *Substantive Condition Schema*: E_1 has the right content or causes S to believe, desire, intend, or do the right things.

Examples of "structural condition theories" include the notable autonomy theories proposed by Frankfurt (1971) and Dworkin (1988). According to those theories, a desire is authentic if the desire-holder identifies with it on a higher level of reflection. To illustrate, Noggle uses the example of an addict who has a first-order desire to use drugs, and a second-order desire to not use drugs: "When a person has both a first-order desire and a second-order desire not to have the first-order desire [...] this repudiated first-order desire is properly regarded as 'a force other than his own'" (p. 89). Thus, in those theories, a desire (element E_1) is authentic if it complies with higher-level desires (E_2). Structural condition theories have been supported in recent writings by, among others, Christman (2009), DeGrazia (2005), and Sjöstrand and Juth (2014).

In "historical condition theories," desires are authentic if they have the right sort of causal history. "The motivating idea behind historical conditions seems to be that a psychological element is authentic if its history is free of the kinds of influences [...] that seem to undermine authenticity" (Noggle 2005, pp. 93–94). Noggle refers to Dworkin, who offers examples of conditions that form the kind of influence that negates authenticity; "hypnotic suggestion, manipulation, coercive persuasion, subliminal influence, and so forth" (1988, p. 18).

Lastly, in contrast to the prior families of theories, "substantive condition theories" are not content-neutral. In substantive condition theories, the content of desires matter to the authenticity of the desires. The following is a hypothetical case which is sometimes used as an example to distinguish between substantive theories and content-neutral theories. Suppose that a woman lives with a man that regularly abuses

her physically and verbally. The woman could choose to leave the man but chooses not to do so. Are the desires underlying woman's choice not to leave the man authentic?

Content-neutral theories of authenticity are concerned with the processes of her choosing. They could conclude that the woman's choice rests on authentic desires. In Noggle's terminology, the woman's decision-making processes could be structurally or historically conditioned so that there is no ground for concluding that her choice builds on inauthentic desires (although this conclusion is improbable).

By contrast, a substantive theory could reach the opposite conclusion, on the grounds that no matter how the woman's decision-making processes are structurally or historically conditioned, the desire to stay with an abusing man cannot be authentic *because* it is the desire to stay with an abusing man; the desire has the wrong content. The reasons why the content is wrong vary between different substantive theories. For instance, one possible explanation is that one cannot authentically desire to fully submit oneself to the wishes of someone else. Submitting oneself fully to others' wishes is to resign as a moral agent, which goes against the very idea of authenticity; one distinguishing factor between inauthenticity and authenticity is that the latter has to do with being self-driven in some sense. The woman cannot authentically choose to submit herself to the man, because one cannot authentically wish to be else-driven.

Content-neutral theories are commonly called "procedural." Procedural theorists hold that a theory of authenticity should be content-neutral mainly because it should not be moralizing or enable undue paternalism. Essentially, it should be content-neutral because it should be morally neutral. Theorists from the substantivist tradition disagree, not least because of the reasons invoked above. The debate between theorists from the two traditions is ongoing (cf. Christman 2004; Oshana 2015), and while the distinction between procedural and substantive theories is relevant in the next section it may be left without further elaboration here.

The taxonomy I have proposed is not of authenticity theories, but of features that various theories share. In the taxonomy, different theories

of authenticity are divided into three categories according to unique features. The categories are *sanctionism*, *originism*, and *coherentism* (Ahlin 2018, pp. 45–47).[2] Here, by "sanctionist theories," for instance, I intend a hypothetical theory which only displays sanctionist features, although the wording is only for pedagogical reasons; authenticity theories can display features from more than one category, and to different degrees. One strength of this taxonomy is that it shows that two distinct families of theories are conflated in Noggle's taxonomy, namely, those that emphasize affirmative self-reflection and those that emphasize coherence. This is elaborated on below. One weakness is that the taxonomy only collects features shared by procedural theories of authenticity.

I call the distinguishing feature of sanctionist theories "affirmative self-reflection." This feature is similar to the structural condition schema in Noggle's taxonomy. Easily put, affirmative self-reflection is to critically scrutinize one's own desires and approve of the result. For instance, suppose that a patient came to know precisely why she has the desire to refuse a medical intervention, reflected critically upon those causes, and concluded that she supports having the desire. The patient would have engaged in affirmative self-reflection and, according to sanctionist theories, her desire to refuse would be authentic.

The distinguishing feature of originist theories is very similar to the historical condition schema in Noggle's taxonomy. In originist theories, desires are inauthentic if they have the wrong sort of origin. One example of an inauthentic desire is one which is "shaped by irrelevant causal factors, by a blind psychic causality operating 'behind the back' of the person" (Elster 1983, p. 16). Elster writes that "desires that have been deliberately chosen, acquired or modified—either by an act of will or by a process of character planning" are authentic (p. 21). That is, desires are authentic if they originate in the right kind of cognitive processes. On another originist account, desires are authentic if they originate in processes of self-discovery and self-definition (Ahlin 2018, p. 46; cf. Meyers 2001, 2005).

[2] Each category can be divided in two classes, namely *cognitivism* and *non-cognitivism*, although these will be left out of the present analysis.

By introducing coherentism, the final category in my taxonomy, a distinguishing feature is fleshed out that is conflated in the structural condition schema in Noggle's taxonomy. In coherentist theories, desires that deviate from the desire-holder's full set of desires are inauthentic. For instance, in Christman's theory of autonomy, a desire is authentic if it is not "alienated" upon self-reflection, "given one's diachronic practical identity and one's position in the world" (2009, p. 155). The notion of "practical identity" should here be understood to mean "a certain pattern of thinking and reacting which, generally speaking, is ours alone; it marks our character and personality" (p. 150). In short, a desire is inauthentic if it does not fit with how the desire-holder's identity has developed over time, and how the identity is presently being sustained. Similarly, Miller argues that an action (here: desire) is inauthentic if it is "unusual or unexpected, relatively important in itself or its consequences, and [has] no apparent or proffered explanation" (1981, p. 24). Thus, in coherentist theories, desires are inauthentic if they are deviating.

Judgments of inauthenticity

The structure of the argument

In this section, I spell out my proposal of what justifies judgments of inauthenticity. In short, the argument has three elements. The first element is a normative thesis determining under which conditions judgments of inauthenticity are justified. It is introduced in the next subsection. The second element is a set of indicators of inauthenticity, i.e., empirical factors that indicate whether the conditions in the first element are met. It is introduced in a subsequent section. The third element, which is also spelled out in an independent subsection, is a clause that delimits the scope of desires and desire-holders which may be justifiably subjected to judgments of inauthenticity. In a final subsection, the elements are collected and formulated as a proposal of what justifies judgments of inauthenticity.

I think of my arguments as contributing to theories in the sanctionist tradition. The tradition is the most influential, and any serious contribution to it should be of interest to autonomy theorists in general.

However, I do not intend to defend sanctionism as such here; that is a different project. This also means that my arguments are intended to be neutral with regard to the content of desires, and thus only concern their procedural forms.

The dissenting self-reflection thesis

It has been pointed out that sanctionist theories suffer from epistemic problems that are difficult to overcome. Sjöstrand and Juth write (2014, p. 21):

> For one thing, it is often difficult to come up with a full explanation as to why we have a certain desire, and even more difficult to make the necessary investigations in order to determine whether or not this explanation is correct.

It may be added that even if this problem is solved, it is also difficult to know whether affirmative self-reflection actually takes place (Ahlin 2018, p. 47):

> [Observing a desire-holder's endorsement of a desire] would require access to advanced (and currently unavailable) neuroimaging technology, in addition to an in-depth knowledge of the psychological nature of endorsement. It would appear that sanctionism is, at the very least, impractical. [It] does not render observable and testable consequences without technology and scientific knowledge yet unheard of, if at all.

However, sanctionism remains a valid and strong theoretical ideal. Consider this thesis, which is formulated with sanctionism as a starting point:

> **The dissenting self-reflection thesis**: Judgments of inauthenticity are justified if there is sufficient reason to believe that the desire-holder would disapprove of having the desire upon informed and critical self-reflection.

In it, affirmative self-reflection is re-stated as a negative. The dissenting self-reflection thesis does not claim to distinguish between authentic and inauthentic desires. The thesis states the conditions under which it is justified to judge that a desire is inauthentic, which means that there could be inauthentic desires that observers for some reason are not justified to call inauthentic. Although the thesis is sanctionist, it is more practical than the theoretical ideal. It facilitates the quest for empirical indicators of inauthenticity. From an observer's epistemically inadequate point of view there are many reasons for a desire-holder to approve of her own desires, while reasons to disapprove of them are fewer, or at least easier to identify. That is, things that indicate inauthenticity are easier to observe than things that indicate authenticity. Therefore, re-stating affirmative self-reflection as a negative has at least one major epistemic merit.

It also has at least one moral merit. The dissenting self-reflection thesis includes a tacit assumption of authenticity; desires should be judged as authentic unless there is evidence of the opposite. There are moral reasons supporting this view. For instance, taking other people seriously, i.e., listening to what they say, respecting their wishes, treating them as "ends in themselves," and so on, seems to require the assumption that they are acting from authentic desires. The dissenting self-reflection thesis complies with those reasons through its tacit assumption of authenticity. Therefore, there is at least one moral merit in building from the dissenting self-reflection thesis rather than from theses of affirmative self-reflection.

The aim of the arguments in what follows is to determine when there is reason to believe that a desire-holder would disapprove of having a desire upon informed and critical self-reflection. The aim is empirical. That is, the aim is to identify empirical factors that indicate that a desire is inauthentic. I have found two possible candidates in the recent literature on authenticity that, when combined, indicate inauthenticity. They are spelled out dialectically after this short but important subsection on a fixed point in the analysis.

A fixed point in the analysis

On the standard account of reflective equilibrium, the analysis allows for "fixed points" that are less sensible to re-evaluation than other matters included in the inquiry (Daniels 2016; Rawls 2001, pp. 29–30). For an illustration, the thesis that it is wrong to torture innocents for mere amusement can be held as a fixed point in an analysis of, for instance, the ethics of war. The present analysis holds one thesis as a fixed point, namely the following.

In one famous case, a 40-year old man developed a sexual interest in children that was later found to be causally connected to a brain tumor (Burns and Swerdlow 2003). When the tumor was removed the pedophilic symptoms disappeared. After some time, the man displayed the same symptoms again, and upon examination it was found that the brain tumor had returned. The causal connection between the man's brain tumor and his sexual desires is clear beyond reasonable doubt. The man went through various medical procedures and a 12-step program for sexual addicts to be able to return to his family and his prepubescent stepdaughter, towards whom he had previously made subtle sexual advances.

Here, for reasons of stable and considered intuitions, I hold this thesis as a fixed point: *It is justified to treat the man's sexual desires as inauthentic*. Holding the thesis as a fixed point does not entail any normative commitments. For instance, it does not mean that I do not hold the man accountable for his actions, or that I would support forced medical interventions aiming to remove the tumor or counteract its causal effects. I merely believe that few or no real cases are better suited to be held as fixed points in the present analysis; if it is not justified to treat the man's sexual desires as inauthentic, it is perhaps never justified to treat any desire as inauthentic.

Indicators of inauthenticity

The originist (or "historical condition") view that desires are authentic if they have the right sort of causal history is intuitively compelling. Desires that are "shaped by irrelevant causal factors" that operate "behind

the back" of the desire-holder seem to be prima facie inauthentic. It is therefore an indicator of inauthenticity:

> **The first indicator of inauthenticity**: It is a reason to believe that a desire-holder would disapprove of having a desire upon informed and critical self-reflection if it is known that the desire is due to causal factors that are not normal to how the desire-holder is otherwise construed, taking both physical and mental dispositions into consideration.

Applying it to the fixed point-case, the indicator provides a plausible explanation for why it is justified to treat the man's sexual desires as inauthentic. However, it requires some elaboration.

Suppose, for the sake of argument, that the man had lived with the brain tumor and its causal influences from birth. He could then have felt a deep and stable connection between his personality and his sexual preferences (disregarding whether he found them morally acceptable) and developed a social identity and a way of life thereafter. He could, for instance, have thought of himself that, "I am the sort of person who cannot live close to playgrounds and schools," and decided to live in solitude, cultivating an interest in botany, rock-climbing, or literature. In this case, the brain tumor is normal to how the man is otherwise construed. It would be less justified to believe that the man would disapprove of his desires upon informed and critical self-reflection. Perhaps it would not be justified at all, as the hypothetical man is known to feel a deep and stable connection between his personality and his sexual preferences. Therefore, the clause in the indicator stating that the causal factors must not be normal to how the desire-holder is otherwise construed is important. The clause also introduces a person-specific quality to judgments of inauthenticity, as what is normal to one person may not be normal to another (which the example in this paragraph illustrates).

However, the first indicator of inauthenticity does not have universal explanatory power. Consider, as a counterexample, a case in which a brain tumor causes a person to have desires that she already has. For instance, a sugar addict may have many reasons to love sugar. Perhaps her parents rewarded her with candy in her early childhood, thus "pro-

gramming" her to have a certain cognitive attitude to sugar, and perhaps she lives next to a chocolate factory and cannot resist the always-present scent of sweets in her neighborhood. Suppose that this person develops a brain tumor causing her to have a desire for sugar. The tumor's causal influences would be just one among many others, and it is therefore not obvious that the sugar addict would disapprove of having those desires upon informed and critical self-reflection. It should be concluded that although the first indicator is reason-giving, it does not by itself provide sufficient reason for judgments of inauthenticity to be justified.

The literature on authenticity includes other possible indicators of inauthenticity. Consider the coherentist view that a desire is inauthentic if it does not fit with the desire-holder's practical identity. It renders a second indicator:

> **The second indicator of inauthenticity**: It is a reason to believe that a desire-holder would disapprove of having a desire upon informed and critical self-reflection if it is known that the desire does not cohere with how the desire-holder's identity has developed over time and is presently being sustained.

Applying the indicator to the fixed point-case, it provides an explanation why it is justified to treat the man's sexual desires as inauthentic. But, it needs to be elaborated.

People in general are rarely fully coherent beings. It is likely that most or all of us have conflicting desires. This is even more certain if we think of humans as intertemporal beings that exist over time; few people have fully coherent desire-sets over time from childhood to old age. Therefore, the second indicator should be understood as pointing at very serious deviations. A hypothetical person's desire to drink beer for lunch and desire to be sober in the afternoon are conflicting, but the conflict is not serious enough for it to be justified to judge any of the two desires as inauthentic. But, in the fixed point-case, the man had deep sexual desires that were at odds with other deep desires, such as that of being part of a loving family and have a normal social life. His sexual desires negated life plans that were important to him. Therefore, they are serious enough to be a reason to believe that a desire is inauthentic.

However, as with the first indicator, the second indicator of inauthenticity lacks universal explanatory power. Consider, for instance, a person who does not suffer from a brain tumor, but who begins to act upon pedophilic desires because of the sudden realization that he has them. That is, suppose that the man's sexual desires were not due to some non-normal causal factor, but that they were latent and emerged upon new stimuli. It would then not matter to judgments of inauthenticity that the desires conflict with the man's practical identity. Therefore, as with the first indicator of inauthenticity, it should be concluded that although the second indicator is reason-giving it does not by itself provide sufficient reason for judgments of inauthenticity to be justified.

The two indicators seem to complement each other. Consider them in combination:

> **A combination of the two indicators of inauthenticity**: There is reason to believe that a desire-holder would disapprove of having a desire to the extent that the desire is known to be due to causal factors that are not normal to how the desire-holder is otherwise construed, taking both physical and mental dispositions into consideration, and to the extent that the desire is known to be incoherent with how the desire-holder's identity has developed over time and is presently being sustained.

The combination explains why it is justified to treat the man's sexual desires as inauthentic in the fixed point-case, while withstanding the counterarguments that have been directed to the two indicators of inauthenticity as separate indicators. It avoids the argument directed to the first indicator regarding new factors that cause people to hold desires that they already have; because they are not deviating, they do not justify judgments of inauthenticity. It also avoids the argument directed to the second indicator regarding normal factors causing incoherent desires; because the factors are not of a certain kind, they do not justify judgments of inauthenticity. Thus, the combination seems to provide a reasonable explanation while withstanding criticism that refute the two indicators of inauthenticity as separate indicators.

However, there are forceful counterexamples also to the combination of the two indicators. Suppose that the man in the fixed point-case already was a pedophile, but that the brain tumor caused his sexual desires in children to disappear. The causal factors would not be normal to how he is otherwise construed, and his new set of desires would be seriously incoherent with his practical identity, yet it seems counterintuitive to conclude that it is justified to believe that the man would disapprove of his new desire-set upon informed and critical self-reflection. On the contrary, he might view the brain tumor as a blessing.

This counterexample is different from the previous; it targets the content of the desires rather than their procedural forms and thus comes from what in Noggle's terminology is called a "substantive condition schema." And, it seems to succeed in one aspect; the combination of the two indicators only seems to justify judgments of inauthenticity when the desires under scrutiny are bad, in some sense. When the desires are good, in some sense, the combination provides counterintuitive conclusions. This can be further illustrated.

In his popular book *The Man Who Mistook His Wife for a Hat*, the neurologist Oliver Sacks reports of a 90-year-old woman who had noticed a "change" (1985, Ch. 11). Around her 88th birthday, she had begun to feel energetic, alive, younger; she had always been shy, but now she flirted with young men, made jokes, and had fun. Her friends thought that her frisky behavior was inappropriate at her age. The woman was feeling extremely well—*too* well—and realized that it could be "Cupid's disease." She had received treatment in her youth, but the infection had only been suppressed, not eradicated. The woman was right, she had neurosyphilis. Upon confirmation of her hypothesis, the woman stated that she did not want to be cured from the infection, as she enjoyed its positive effects, but that she did not want it to get worse either.[3] Contrary to what the combination of the two indicators would suggest, the woman did not disapprove of her desires upon informed and critical self-reflection, in spite of the fact that they were both alien and deviating.

[3] Sacks eventually found a treatment that conformed to the woman's wishes.

Thus, the combination seems to justify judgments of inauthenticity when the desires under scrutiny are bad in some sense, but not when they are good in some sense.[4] This observation needs to be sorted out and answered.

A delimiting clause

The dialectic in the above subsection builds on theories and propositions from the procedural tradition of autonomy theorizing. Yet, the final counterexample which convincingly refutes the combination of the two indicators of inauthenticity seems to be substantive. The example succeeds in showing that the combination only renders plausible conclusions regarding desires with a certain kind of content. Thus, the dialectic appears to rest on tacit substantive assumptions that must be made explicit and explained.

In the fixed point-case, the assumption is that it would be better for the man to not have those sexual desires. It would be better according to objective moral standards, as not being a pedophile is morally better than being one. The clause could be added to the combination of the two indicators, "unless the desire is better according to objective moral standards." Judgments of when there is sufficient reason to believe that a desire-holder would disapprove of a desire would then be normative. Such moralizations are precisely what proceduralists wish to avoid. The substantive assumption that must be made explicit can therefore not be objective in the sense reflected by this line of thought.

It would also be better for the man to not have those sexual desires according to his own subjective moral standards. It can reasonably be assumed that the man did not want to have those desires, as he underwent a 12-step program for sexual addicts in addition to the various medical procedures to be able to return to his family. A different clause could thus be added to the combination, "unless the desire is better according to subjective moral standards." However, judgments of inauthenticity

[4] One plausible line of thought is instead that judgments of inauthenticity may be justified in either case, but that they are only interesting when the desire under scrutiny is bad in some sense. However, mainly for reasons of space, this possibility is not further explored here.

would then be self-referential and non-guiding. The full theory would essentially carry the meaning, "there is reason to believe that the desire-holder would disapprove of having the desire if it is new and deviating, unless the desire-holder approves of having it," or simply, "the desire-holder would disapprove of having the desire unless she approves of having it." Therefore, the substantive assumption that must be made explicit cannot be subjective in the sense reflected by this line of thought either.

However, there is one possible middle-way between those two lines of thought. The following delimiting clause can be added:

> **Delimiting clause**: Here, judgments of inauthenticity only target desire-holders who are known to carry a general wish to live according to the prevailing social and moral standards, and desires that are seriously undesirable according to those standards.

The man in the fixed point-case is such a desire-holder, which is known from his efforts to defeat the symptoms of his brain tumor, and his desires are undesirable accordingly, which is known from observations of the prevailing social and moral standards. More fully spelled out, the judgment is then that there is reason to believe that the man in the fixed point-case would disapprove of having his desires as they are alien, incoherent, and undesirable according to the standards which it is known that he wishes to follow. The delimiting clause also goes well with the old woman whose syphilis caused her to have alien and deviating desires. Her frisky mood and behavior is not undesirable enough either socially or morally. Therefore, the woman's desires are not of the kind that may justifiably be targeted by judgments of inauthenticity.

I expect four immediate objections to the delimiting clause. Each is due to the fact that it brings normative content into judgments of inauthenticity.

First, proceduralist protests can be expected. Proceduralists hold that judgments of inauthenticity should be content-neutral because they should be morally neutral. However, they should not worry. There is no risk for undue paternalism, as the clause also states that judgments of inauthenticity here only target people who are known to wish to live according to the same normative content that the judgment builds on. And, the normative content does not concern the distinction between

authentic and inauthentic desires, but when it is justified to treat desires as inauthentic, i.e., an inherently normative problem. There must be moralization, and proceduralists should grant that applying the desire-holder's own moral standards is at least less problematic than applying someone else's.

Second, it may be objected that adding the delimiting clause to the theory entails that judgments of inauthenticity would be self-referential and non-guiding, in the sense discussed above. The full theory would essentially carry the meaning, "the desire-holder would disapprove of having the desire unless she approves of having it." The objection obscures an important difference between the two additions. In the above discussion, the worry of circularity is due to the fact that the addition to the combination reads, "unless the desire is better according to subjective moral standards." In that addition, the desire-holder's attitude to the specific desire which is under scrutiny is known and added to the analysis. In the delimiting clause, on the contrary, nothing is said about any specific desire. Instead, the delimiting clause adds the desire-holder's attitude toward the prevailing social and moral standards (according to which specific desires may be good or bad). Thereby, the addition is substantive enough to be analytically potent, while being sufficiently content-neutral to avoid being problematically circular.

Third, substantivist protests can also be expected. Substantivists hold that the content of a desire should be of a certain kind for the desire to be authentic, and the normative content which is here brought into judgments of inauthenticity is not objective. If the prevailing moral standards are wicked, that wickedness is brought into the judgment and is thus acted upon—surely, the substantivist might say, it must be wrong in and by itself to act upon wicked moral standards. However, the task is here to justify judgments of inauthenticity in light of the desire-holder's values. In all other matters, it is a value-neutral project. The current project aims to contribute to the justification of judgments of the possible inauthenticity of desire-holders' wicked views, not the justification of their views as such.

Fourth, delimiting the analytical scope to desire-holders who are known to wish to live according to the prevailing social and moral standards may entail that a lot of people are left out. The current analysis does not provide guidance with regard to people whose values and motivational sets are unknown or deviating from the prevailing social and moral standards. My answer is simply that including those people is not morally justified, precisely because their values and motivational sets are unknown or deviating. The present theory ultimately aims to be proceduralist and thus both non-moralizing and anti-paternalist. Therefore, it is not a problem that the analytical scope is narrow in this sense. On the contrary, this narrowness counts in favor of the theory as a whole.

Justifying judgments of inauthenticity

Adding the delimiting clause to the dissenting self-reflection thesis and to the combination of the two indicators of inauthenticity results in the following:

> (1) Here, judgments of inauthenticity only target desire-holders who are known to carry a general wish to live according to the prevailing social and moral standards, and desires that are seriously undesirable according to those standards.
>
> (2) Judgments of inauthenticity are justified if there is sufficient reason to believe that the desire-holder would disapprove of having the desire upon informed and critical self-reflection.
>
> (3) There is reason to believe that a desire-holder would disapprove of having a desire to the extent that it is known to be due to causal factors that are not normal to how the desire-holder is otherwise construed, taking both physical and mental dispositions into consideration, and to the extent that it is known to be incoherent with how the desire-holder's identity has developed over time and is presently being sustained.

1 through 3 justifies judgments of inauthenticity. In shorter terms, and in light of the arguments above, the justification may read:

> For persons who wish to live according to the prevailing social and moral standards and desires that are seriously undesirable according to those standards, it is justified to judge that a desire is inauthentic to the extent that it is due to causal factors that are alien to the person and to the extent that it deviates from the person's practical identity.

Concluding remarks

To conclude, it is not clear how the notion of authenticity should be understood, nor what its place is in the contemporary autonomy-oriented bioethical paradigm. However, it is both common and reasonable to treat the notion as concerning desires. As such, it is a problem to determine what justifies judgments of inauthenticity. A content-neutral solution to that problem has been proposed building from the sanctionist tradition (a starting point and a tradition which both require elaborate and independent defenses).

I do not claim that my proposal is the only way to justify judgments of inauthenticity. Instead, it should be understood as generally reason-giving in a larger framework of reflective equilibrium, and applying it requires substantial moral deliberation. For instance, very little has been said about what it means that a desire is "seriously" deviating from a person's practical identity. Further guidance in that particular matter is found in, e.g., Christman (2009, pp. 149–156). Likewise, 1 through 3 and the summarized proposal are expressed in terms of degrees rather than in necessary or sufficient conditions. Therefore, it must always be a matter of deliberation to determine, e.g., to which extent a desire is known to be alien to the desire-holder. Further guidance in how to apply the proposal in practice may be found in the methodological discussions in Beauchamp and Childress (2013) and Beauchamp and Rauprich (2016).

Furthermore, it must be noted that decisions are not autonomous only because it is not justified to judge that the underlying desires are

inauthentic. Although the desires must be treated as authentic, the desire-holder may, e.g., be incapable of realizing and assessing the implications of the decision. In short, authentic decisions do not necessarily amount to autonomous decisions.

I have here attempted to bridge the gap between theoretical ideals of authenticity and real authenticity-related problems in practical biomedical settings. Future contributions to authenticity theory may determine whether I have succeeded in that aim.

References

Ahlin, J. (2018). The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent. *Medicine, Health Care and Philosophy* 21(1), 43–50.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5 ed.). Washington, DC: APA.

Bauer, K. (2017). To Be or Not To Be Authentic. In Defence of Authenticity as an Ethical Ideal. *Ethical Theory and Moral Practice* 20(3), 567–580.

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of Biomedical Ethics* (7 ed.). Oxford University Press.

Beauchamp, T. L., & Rauprich, O. (2016). Principlism. In H. ten Have (Ed.), *Encyclopedia of Global Bioethics* (pp. 2282–2293). Springer.

Burns, J. M., & Swerdlow, R. H. (2003). Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign. *Archives of Neurology* 60(3), 437–440.

Christman, J. (2004). Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves. *Philosophical Studies* 117(1), 143–164.

—. (2009). *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge University Press.

Daniels, N. (2016). Reflective Equilibrium. *Stanford Encyclopedia of Philosophy*. Retrieved (October 10, 2018) from https://plato.stanford.edu/archives/fall2018/entries/reflective-equilibrium/

DeGrazia, D. (2005). Human Identity and Bioethics. Cambridge University Press.

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.

Eyal, N. (2012). Informed Consent. *Stanford Encyclopedia of Philosophy*. Retrieved (January 9, 2019) from http://plato.stanford.edu/entries/informed-consent/

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1), 5–20.

Grisso, T., Appelbaum, P. S., & Hill-Fotouhi, C. (1997). The MacCAT-T: A Clinical Tool to Assess Patients' Capacities to Make Treatment Decisions. *Psychiatric Services* 48(11), 1415–1419.

Juth, N. (2005). *Genetic Information: Values and Rights. The Morality of Presymptomatic Genetic Testing*. Acta Universitatis Gothoburgensis.

Lester, R. J. (2009). Brokering Authenticity Borderline Personality Disorder and the Ethics of Care in an American Eating Disorder Clinic. *Current Anthropology* 50(3), 281–302.

Meyers, D. T. (2001). Authenticity for Real People. In B. Elevitch (Ed.), *Proceedings of the Twentieth World Congress of Philosophy: Philosophy of Mind and Philosophy of Psychology* (pp. 195–202). Philosophy Documentation Center.

—. (2005). Decentralizing Autonomy: Five Faces of Selfhood. In J. Christman & J. Anderson (Eds.), *Autonomy and the Challenges to Liberalism* (pp. 27–55). Cambridge University Press.

Miller, B. L. (1981). Autonomy & the Refusal of Lifesaving Treatment. *The Hastings Center Report* 11(4), 22–28.

Nelson, R. M., Beauchamp, T. L., Miller, V. A., Reynolds, W., Ittenbach, R. F., & Luce, M. F. (2011). The Concept of Voluntary Consent. *The American Journal of Bioethics* 11(8), 6–16.

Noggle, R. (2005). Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (pp. 87–108). Cambridge University Press.

Oshana, M. A. L. (2015). Is Social-Relational Autonomy a Plausible Ideal? In M. A. L Oshana (Ed.), *Personal Autonomy and Social Oppression: Philosophical Perspectives* (pp. 3–24). Routledge.

Rawls, J. (2001). *Justice as Fairness: A Restatement* (E. Kelly Ed.). The Belknap Press of Harvard University Press.

Sacks, O. (1985). *The Man Who Mistook His Wife For a Hat: And Other Clinical Tales*. Touchstone Books.

Sjöstrand, M., & Juth, N. (2014). Authenticity and Psychiatric Disorder: Does Autonomy of Personal Preferences Matter? *Medicine, Health Care and Philosophy* 17(1), 115–122.

Tan, D. J. O. A., Hope, P. T., Stewart, D. A., & Fitzpatrick, P. R. (2006). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. *Philosophy, Psychiatry, & Psychology: PPP* 13(4), 267–282.

Taylor, C. (1991). *The Ethics of Authenticity*. Harvard University Press.

Taylor, J. S. (2005a). *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy*. Cambridge University Press.

—. (2005b). Introduction. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy* (pp. 1–29). Cambridge University Press.

White, L. (2018). The Need for Authenticity-Based Autonomy in Medical Ethics. *HealthCare Ethics Committee Forum* 30(3), 191–209.

 The article is originally published as: Ahlin, J. (2018). What Justifies Judgments of Inauthenticity? *HealthCare Ethics Committee Forum* 30(4), 361–377. **The following changes have been made in the present version:** (1) The list of references has been updated (a) according to the reference standards of this thesis and (b) to reflect that White (2018) has been assigned to an issue of *HealthCare Ethics Committee Forum* since the publication of this article. (2) The word "for" has been removed from between the words "explanation" and "why" on p. 78.

A Non-Ideal Authenticity-Based Conceptualization of Personal Autonomy

Abstract: Respect for autonomy is a central moral principle in bioethics. The concept of autonomy can be construed in various ways. Under the non-ideal conceptualization proposed by Beauchamp and Childress, everyday choices of generally competent persons are autonomous to the extent that they are intentional and are made with understanding and without controlling influences. It is sometimes suggested that authenticity is important to personal autonomy, so that inauthenticity prevents otherwise autonomous persons from making autonomous decisions. Building from Beauchamp and Childress's theory, this article develops a non-ideal authenticity-based conceptualization of personal autonomy. Factors that indicate inauthentic decision-making are explicated, and the full concept is defended from three expected objections. The theory is then tested on a paradigm case which has concerned theorists and practitioners for some time, namely the possible inauthenticity of anorexia nervosa patients' decision-making. It is concluded that the theory seems to be fruitful in analyses of the degree of autonomy of patients' decision-making, and that it succeeds in providing reliable action-guidance in practical contexts.

Keywords: Autonomy, authenticity, anorexia nervosa, healthcare, bioethics

Introduction

Respect for autonomy is a central moral principle in bioethics. The concept of autonomy can be construed in various ways. Under Beauchamp and Childress's non-ideal conceptualization, everyday choices of generally competent persons are autonomous to the extent that they are intentional and are made with understanding and without controlling in-

fluences (2013, p. 104ff). It is sometimes suggested that authenticity, i.e., being "real," "genuine," "true to oneself," or similar, is important to personal autonomy, so that inauthenticity prevents otherwise autonomous persons from making autonomous choices. Yet, while the notion has previously been included in ideal conceptualizations of autonomy, there have at least to my knowledge not been any attempts at incorporating authenticity in a non-ideal conceptualization of personal autonomy.[1]

Elsewhere, I have proposed that judgments of inauthenticity in others are justified under certain conditions (Ahlin 2018b). In this article, I adjust those conditions for the present purposes and add them to Beauchamp and Childress's account of autonomy. The result is a non-ideal authenticity-based conceptualization of autonomy supplemented with relatively easy detected factors that indicate non-autonomous decision-making.

The article is structured as follows. First, I account for and briefly discuss White's recently proposed ideal account of authenticity-based personal autonomy. This is followed by a more elaborate explication of Beauchamp and Childress's non-ideal account. In the subsequent section, I introduce the conditions under which judgments of inauthenticity are justified and add them to Beauchamp and Childress's account of autonomy to render a non-ideal authenticity-based conceptualization of autonomy. Factors that indicate non-autonomous decision-making are explicated, and three expected objections are met. Then, I apply the complete account to a case which has been thoroughly discussed in the literature on authenticity, namely a patient who suffers from anorexia nervosa and expresses potentially distressing wishes concerning her own medical situation. A brief final section concludes the discussion.

[1] One possible exception is Swanson, who argues that a non-ideal conceptualization of autonomy should include "a consistent sense of personal identity" (2017, pp. 51–3).

Personal autonomy

Ideal accounts of authenticity-based autonomy

There are many different usages of the terms "ideal theory" and "non-ideal theory" (Valentini 2012). Here, I intend "ideal theory" to designate some model—in this case of autonomy and/or authenticity—that is largely hypothetical. Few or no persons or decisions are ever fully autonomous or authentic in this sense, as the conditions under which ideal autonomy or authenticity obtains are perfect or conceptual. By "non-ideal theory," I intend accounts that are not constructed accordingly. The approach is sometimes also known as "realist" or "problem-oriented," as it starts from actual people, facts, conditions, etc., in the real world rather than in some theoretical model.

There are various theories aiming to explain authenticity, none of which takes precedence over others (Ahlin 2018a; Noggle 2005). The relevant problem in practical biomedical contexts is not to determine what is authentic, but what is inauthentic. One example of when the notion has been invoked in biomedical contexts is when patients suffering from anorexia nervosa have said that they would rather die than gain weight (Tan et al. 2006). The claim appears to be inauthentic and, arguably, for that reason also non-autonomous (see, e.g., Sjöstrand and Juth 2014). This is discussed at greater length in a below section.

In one of the most recent contributions to authenticity theory in biomedical ethics, White (2018) argues in favor of using the notion of authenticity as a frame of reference in assessments of the validity of patients' healthcare decisions. More specifically, it should be used in accounts of autonomy as a means to protect high-stake choices from being overridden. White suggests that the notion of authenticity should provide an underlying frame of reference that allows assessments of whether a particular healthcare decision is adequately understood or appreciated. The theory of authenticity which White adopts is broadly Lockean. In it, authenticity concerns the "self," which should be understood as a set of "enduring, stable overlapping psychological elements, including values, beliefs and desires" (pp. 193–194). This entails that the

validity of a patient's healthcare decisions should be assessed in relation to the historical construction of the patient's self. A healthcare decision that conflicts with the psychological elements that constitute the patient's "self" is inauthentic, and this inauthenticity should influence assessments mainly of the patient's decision-making competence.

Other, less recent, similar ideal theories include Christman (2009) and Juth (2005). Christman argues that "Autonomy involves competence and authenticity; authenticity involves non-alienation upon (historically sensitive, adequate) self-reflection, given one's diachronic practical identity and one's position in the world" (2009, p. 155). Juth offers this minimalist definition of personal autonomy: "A person, in a situation, is autonomous to the extent that she does what she decides to do, because she decides to do it, and decides to do what she wants to do, because she wants to do it" (2005, p. 137). He proceeds to argue that, in this analysis, authenticity is one of three components of autonomy (the other two are *decision competence* and *efficiency*) (ibid).

These suggestions are ideal. White considers some of the practical restrictions in healthcare contexts, such as the epistemic difficulties of determining inauthenticity in others, but builds from a theoretical model rather than from real patients in real contexts. I do not claim that these constructs are mistaken or irrelevant, but I wish to propose a non-ideal alternative to them. In contrast to White's theory, my proposal is to add the notion of authenticity to a set of conditions which both together and separately indicate that a patient's decision-making is non-autonomous (in different aspects). This furthers the theoretical approach to autonomy theory which takes authenticity as one of the basic conditions of autonomy, but differs from previous contributions in that it is non-ideal rather than ideal.

Beauchamp and Childress's non-ideal account of autonomy

In their book *Principles of Biomedical Ethics* (2013), Beauchamp and Childress propose a non-ideal conceptualization of autonomy. Their account builds on the premise that everyday choices of generally competent persons are autonomous (p. 104). Autonomous actions are then

analyzed "in terms of normal choosers who act (1) intentionally, (2) with understanding, and (3) without controlling influences that determine their action" (ibid). Essentially, the current project is to add a fourth condition to that analysis, namely authenticity. First however, the conditions just mentioned must be further elaborated.

The premise that everyday choices of generally competent persons are autonomous includes standards of incompetence, i.e., conditions that negate a person's decision-making capabilities. Beauchamp and Childress suggest seven types of related inabilities, including the inability to express or communicate a choice, the inability to understand one's situation and its consequences, and the inability to understand relevant information (2013, p. 118). These mark a threshold level of decision-making competence, so that persons who display one or more inabilities should be judged as less competent or incompetent to make the decision in question.

The condition of intentional action is explicated through a contrast with accidental action. Acting intentionally requires a plan, i.e., a "representation of the series of events proposed for the execution of [the] action" (p. 104). Accidental actions are not planned accordingly. Intentional actions "correspond to the actor's conception of the act in question," whereas accidental actions do not (ibid).

The condition of understanding means that an act is non-autonomous if the agent does not adequately understand it (ibid). Having an adequate understanding is different from having a full understanding. For illustration, consider the so-called "butterfly effect," i.e., a common term designating the fact that even small interventions in a system may have significant effects on an aggregated scale. For instance, flicking a cigarette butt in the dry woods is a small act that may lead to a huge wildfire and thousands of people having to relocate. On Beauchamp and Childress's account, an agent is not required to have a full understanding of the "butterfly effect" of an act for it to be autonomous. It should not be said that the person flicking the cigarette butt acted non-autonomously because she did not know that the act would have those significant effects. Adequate understanding, i.e., a reasonable estimation of the nature, meaning, and outcome of the act in question, suffices for it to be autonomous.

Instead of spelling out precisely what "adequate" understanding means, Beauchamp and Childress mention factors that may limit understanding, such as "illness, irrationality, and immaturity" (ibid). When the condition of understanding concerns a person's informed consent to treatment, Beauchamp and Childress list "the nature and purpose of the intervention, alternatives, risks and benefits, and recommendations" as "typically [...] essential" (p. 132). Thus, having an adequate understanding of an act involves awareness of relevant and reasonably foreseeable facts that are central to the act in question. Most importantly, for an act to be autonomous the agent must understand the basics of how it is likely to affect her own person and her way of life.

Finally, the third condition concerns acting without controlling influences (pp. 104–105). Controlling influences may be external to the agent, such as when she is coerced or manipulated into performing some act, or internal to her, such as when she is drunk or suffers from some mental disorder. Obviously, human beings are almost always subjected to some controlling influence. That comes with being a social animal. It is natural to us to lead our lives after the expectation of others, at least to some extent, and our expectations of ourselves are certainly at least partly socially constructed. It is likely that no human being has ever been completely free from controlling influences.

But, Beauchamp and Childress note that controlling influences, unlike the binary notion of intentional and unintentional actions, come in degrees (p. 105). Influences such as coercion and manipulation are controlling to a greater extent than, for instance, the social expectations that women should be beautiful and men should be strong. Because they are more controlling, coercion and manipulation have a greater effect on the autonomy with which an agent acts. Likewise, internal influences such as severe drug addiction may have a greater effect on the autonomy of a person than, for instance, socially contingent self-constraints. Thus, considering only the third and final condition in Beauchamp and Childress's account of autonomy, an act is autonomous to the extent that it is free from controlling influences.

It should be noted that Beauchamp and Childress adhere to a theory of justification and methodology that builds from John Rawls's theory of reflective equilibrium (Beauchamp and Childress 2013, pp. 390–429). On my understanding, they hold that a normative claim is justified to the extent that it is coherent with other relevant claims in moral and factual matters, and with our stable and considered intuitions regarding the problem in question. Elsewhere, Beauchamp writes that their method aims to produce "coherent strings of norms that connect basic principles, derivative norms, and context-specific judgments" (Beauchamp and Rauprich 2016, p. 6). In what follows, my arguments should be understood in light of this methodological approach. I return briefly to these methodological comments below in a discussion about so-called "underdetermined" moral concepts.

Authenticity as a condition of autonomous choosing

Justifying judgments of inauthenticity

In my proposal, judgments of inauthenticity in others concern their decision-making, or more precisely their desires, as desires are the most basic element in ordinary preference-forming and, thus, in decision-making. I follow Taylor (2005), Sjöstrand and Juth (2014), and others in this desire-oriented approach. For reasons of justification, judgments of inauthenticity are delimited to concern only a certain kind of persons, namely those whose medical condition may influence their decision-making so that they hurt themselves or others. Examples of such persons include an anorexia nervosa patient who expresses a wish to die rather than gain weight and someone with a brain tumor that causes him or her to develop pedophilic sexual desires.[2] For those persons and the desires underlying their healthcare decisions, I argue that it is justified to judge that a desire is inauthentic to the extent that it is due to causal

[2] The example with the pedophilic desires is not hypothetical; this happened to an otherwise normally functioning adult man. See Burns and Swerdlow (2003) and the below.

factors that are alien to the person and to the extent that it deviates from the person's practical identity.[3]

Thus, there are three main elements in my proposal that require some elaboration here; the kind of people and desires included in the analysis, the notion of alien causal factors, and the notion of deviation. For pedagogical reasons, I will go through them in the opposite order. However, first I present some brief notes on the theoretical foundation of my proposal.

In one tradition, the distinguishing feature between authentic and inauthentic desires is whether the desire-holder would endorse her own desires upon critical and informed self-reflection. The tradition is known mainly from Frankfurt (1971) and Dworkin (1988) and has been supported more recently by Juth (2005) and DeGrazia (2005), among others. It has been noted that the distinguishing feature is difficult to observe in others (see, e.g., Ahlin 2018a; Sjöstrand and Juth 2014; and; Swindell 2009). That is, it is difficult to know whether a person would endorse her own desires upon informed and critical self-reflection, and therefore the theoretical ideal is impractical, at best. However, the distinguishing feature can be reversed, so that a desire is inauthentic if the desire-holder would *disapprove* of having it upon critical and informed self-reflection (cf. Juth 2005, p. 153). Then, it is less difficult to observe inauthenticity in others, as empirical factors that indicate that a desire-holder would in fact disapprove accordingly can be identified and articulated in detail. The theory which is presently being spelled out builds on this reversed version of the Frankfurt–Dworkean ideal. The notions of alien causal factors and of deviation are empirical factors indicating that a desire-holder would disapprove of having a desire upon critical and informed self-reflection.

That is the theoretical foundation of my proposal. It is an ideal theory of what distinguishes authenticity from inauthenticity. What

[3] This proposal builds on Ahlin (2018b) but includes other kinds of persons and desires. In Ahlin (2018b), persons that are known to wish to adhere to the prevailing social and moral standards and desires that are seriously undesirable according to those standards are justifiably targeted by judgments of inauthenticity. I will not elaborate further on this difference here.

follows here, however, does not concern that distinction per se, but the justification of judgments of inauthenticity, i.e., what justifies the judgment that a desire meets the conditions for being inauthentic. That justification is phrased in non-ideal terms, and builds from empirical factors in real persons and contexts.

Consider a person who suddenly displays a desire that is seriously deviating from her practical identity, i.e., the way she usually thinks, behaves, and functions socially (cf. Christman 2009, pp. 149–156). One hypothetical example is Anna, a professional ballet dancer known to love dancing more than anything else, who after being injured refuses to undergo a minor treatment that would enable her to continue dancing (Ahlin 2018a, p. 44). Her refusal builds on desires that are seriously deviating from her practical identity. Therefore, the case invites the thought that Anna's desires are inauthentic. However, the judgment is not justified. The reasons for why Anna makes the surprising decision to refuse treatment is unknown. Because we do not know the causal history of her desires, we are not justified in making the judgment that they are inauthentic.

Now, consider a 40-year old man who suddenly developed a sexual interest in children that was causally connected to a brain tumor (Burns and Swerdlow 2003). When the tumor was removed the pedophilic symptoms disappeared, and when the symptoms later returned it was found that the brain tumor had grown back. There is no doubt that the tumor caused the man's sexual interests. Thus, the causal factors of the man's desires were alien to how he was otherwise construed, which intuitively seems to justify the judgment that they are inauthentic. However, alien causal factors do not suffice to justify that judgment. For instance, sometimes alien empirical factors cause non-alien desires, such as the hypothetical case of a sugar addict who develops a brain tumor causing cravings for sweets. Therefore, it is not justified to make the judgment that an alien desire is inauthentic merely because of its causal history.

The two notions seem to do well when combined, so that desires that are both deviating from the desire-holder's practical identity and are due to alien causes indicate inauthenticity. However, there is one major flaw in the suggestion that the combination would justify judgments of

inauthenticity. Consider a person who displays a deviating and alien desire which is good, in some sense. For instance, in one case, a 90-year old woman who was otherwise quiet and shy suddenly started to make jokes and flirt with young men (Sacks 1985, chap. 11). Her "frisky behavior" was found to be due to untreated syphilis, i.e., an alien cause of deviating desires. But, the woman enjoyed her new self and did not want it to go away. It appears to be unjustified to judge that her desires are inauthentic, in spite of the fact that they are both deviating and due to alien causes.[4]

Therefore, in my proposal, judgments of inauthenticity should be delimited to persons whose medical condition may influence their decision-making so that they hurt themselves or others. Then, it is justified to make the judgment that, for instance, the 40-year old man who developed a sexual interest in children had inauthentic desires while it is not justified to make the judgment that the 90-year old woman who developed a new way of life had inauthentic desires.

Thus, to summarize my proposal:

> For persons whose medical condition may influence their decision-making so that they hurt themselves or others, it is justified to judge that an underlying desire is inauthentic to the extent that it is due to causal factors that are alien to the person and to the extent that it deviates from the person's practical identity.

In the next subsection, I will incorporate it in Beauchamp and Childress's account of personal autonomy.

A non-ideal authenticity-based conceptualization of autonomy

The basic premise in the theory is that everyday choices of generally competent persons are autonomous. Call such persons "normal." A second basic premise is now added: choices made by otherwise normal persons who suffer from some medical condition that may influence their decisions so that they hurt themselves or others are sometimes inauthentic.

[4] It may also be noted that desires can be inauthentic although the available empirical evidence does not suffice for observers to be justified in making that judgment.

Thus, the theory has two basic premises with different functions. The first premise is directly connected to the autonomy of persons; a person who is not generally competent according to the standards of incompetence elaborated on above is less autonomous than a person who is generally competent. However, a person who suffers from some medical condition of the kind discussed here is not necessarily less autonomous than one who does not suffer from such conditions; for instance, the medical condition may not *actually* influence her decisions merely because it can *potentially* do so. Thus, the second premise is only indirectly connected to the autonomy of persons; it enables judgments of inauthenticity through conditions that will be spelled out shortly.

The fourth condition of authenticity may now be added to Beauchamp and Childress's list, for stylistic reasons between the first and second condition. Thus, autonomous choice can be analyzed in terms of normal persons who act (1) intentionally, (2) from authentic desires, (3) with understanding, and (4) without controlling influences that determine their action. Each condition is assumed to apply until there is reason to believe otherwise. That is, normal persons are assumed to act, e.g., intentionally or with understanding unless something indicates the opposite. It remains here to spell out in practically usable terms factors that indicate that an otherwise normal person acts from inauthentic desires.

Two factors indicate inauthenticity. Both must be present for a judgment of inauthenticity to be justified. In practically useful terms, they read:

> **The factor of deviation** It is a factor indicating inauthenticity that the desire under scrutiny does not cohere with how the desire-holder's identity has developed over time and is presently being sustained.
>
> **The factor of alien causes** It is a factor indicating inauthenticity that the desire under scrutiny is due to causes that are not normal to how the desire-holder is otherwise construed, taking both physical and mental dispositions into consideration.

It should be noted that the factors come in degrees and are sensitive to judgment. For instance, a desire may deviate from a person's practical identity, but only insignificantly. To illustrate, Anna, the hypothetical professional ballet dancer, may have a desire to drink beer on the evening before an important show. The desire conflicts with her desires to stay focused and do everything that is in her power to perform well on the show, although the deviation from Anna's desire-set is not significant enough to indicate inauthenticity. That is, Anna may have the authentic desire to drink beer on the evening before an important show. Deviations should be more serious than that to merit the judgment that a desire is inauthentic. Had Anna instead had the desire to try heroin, the judgment may have been different due to the seriousness of the deviation.

Furthermore, as explained above, in this framework judgments of inauthenticity are only justified regarding a certain kind of persons. Therefore, justification of such judgments requires knowledge of the person's medical condition and substantive deliberation on whether the person may hurt themselves or others. Thus, judgments of inauthenticity are a matter of practical and context-sensitive deliberation in particular cases. Thereby, the present proposal—as Beauchamp and Childress's original account of autonomy—is conceptually underdetermined. It is a structure for rational deliberation on the authenticity of decisions made by otherwise normal persons, but it does not include complete specifications of how the involved concepts apply in particular cases and contexts. As such, the proposal should be understood not in terms of, e.g., necessary and sufficient conditions, but as generally reason-giving in a framework of reflective equilibrium.[5]

Objections

In this subsection, I respond to three (internally independent) objections to my proposal; the threshold for making a judgment of inauthenticity appears to be too high, the condition of authenticity brings normative

[5] For a more in-depth discussion of how underdetermined moral concepts should be applied in practical contexts, see Beauchamp and Rauprich (2016).

content into an otherwise value-neutral conceptualization of autonomy, and, finally, my suggested addition to Beauchamp and Childress's concept of autonomy is superfluous.

Because both factors that indicate inauthenticity must be met for a judgment of inauthenticity to be justified, it appears that few actions would ever be judged as inauthentic. This is not a bad thing; the conceptualization of autonomy which is defended here is ultimately anti-paternalist. From an anti-paternalist perspective, it is good that most actions are treated as autonomous and that factors that indicate the opposite are few. What is important is instead that the conceptualization is accurate.

Furthermore, the factors do in fact support judgments of inauthenticity in real cases (see the next section). Consider, for instance, patients suffering from borderline personality disorder (BPD). Some BPD patients are characterized by unstable "selves" and may, for instance, display sudden and dramatic shifts in goals, values, vocational aspirations, types of friends, and so on (Lester 2009). In generic cases, both factors indicating inauthenticity are thus present; BPD patients are otherwise normal persons with seriously deviating desires that are due to alien causes. Their actions and healthcare decisions are non-autonomous, and the present non-ideal conceptualization of autonomy enables the reliable judgment that they are non-autonomous for authenticity-related reasons.

Proceeding with the second objection, it is true that the condition of authenticity brings normative content into the conceptualization of autonomy through the second basic premise, i.e., that choices made by otherwise normal persons who suffer from some medical condition that may influence their decisions so that they hurt themselves or others are sometimes inauthentic. But, the concept was never value-neutral. Most importantly, Beauchamp and Childress's standards of incompetence are value-laden (2013, pp. 114–20). To paraphrase Buchanan and Brock, the proper standard of incompetence must be chosen; it cannot be discovered (1990, p. 47). Choosing such standards involves moral assessment and deliberation. Thus, any judgement that a person is incompetent to make a certain healthcare decision is moralizing, because the standards of incompetence are morally loaded.

Finally, the third objection is that my suggested addition to Beauchamp and Childress' concept of autonomy is superfluous; their concept already accounts for concerns of inauthenticity through the condition of controlling influences.[6] As explained above, Beauchamp and Childress analyze autonomous actions in terms of normal choosers who act "without controlling influences that determine their action" (2013, p. 104). Some controlling influences are internal to the agent, such as, e.g., psychiatric disorders and drug addiction (p. 138). Therefore, the argument goes, as my suggestion builds on the notion of alien causal factors—understood as internally controlling influences—it adds nothing substantial to Beauchamp and Childress's concept.

However, although they mention the possibility of internally controlling influences, Beauchamp and Childress do not focus on them in their conceptualization of autonomy (pp. 104–105, 138). In fact, internally controlling influences are almost completely left out of the discussion of the condition of non-control. Thus, I see my suggested addition as a contribution to Beauchamp and Childress's concept as it explicates one kind of internally controlling influence. It enables analysis in one instance of non-control that was previously theoretically underdeveloped. Furthermore, the addition suggests that this kind of internally controlling influence should be understood in terms of authenticity specifically, and not in other terms. Thereby, the addition also connects one kind of internally controlling influences to an already established theoretical school of thought, namely the Frankfurt–Dworkean.

Thus, the three objections that the threshold for making a judgment of inauthenticity is too high, that the condition of authenticity brings normative content into an otherwise value-neutral concept, and that my suggested addition is superfluous do not overthrow my proposed authenticity-based conceptualization of autonomy.

[6] I am grateful to an anonymous reviewer for pointing this out.

Testing the theory

Some methodological remarks

A non-ideal account of autonomy is good only insofar as it provides real normative guidance in practical contexts. Therefore, in this section, I apply the account in an analysis of a healthcare decision made by a person suffering from anorexia nervosa. The person declined medical treatment. The test consists in analyzing whether the desires underlying that decision were inauthentic.

Because of its non-ideal nature, it does not suffice to test the theory on a generic case-description of anorexia nervosa; real testing requires a real case. However, there are no in-depth individual case-descriptions focusing on anorexia nervosa in the bioethical literature on authenticity. Therefore, I have here constructed a hypothetical case building from two interview studies conducted with anorexia nervosa patients, namely Hope et al. (2011) and Tan et al. (2006). The studies have been influential in the bioethical debate on authenticity and are generally considered to be authoritative in this context. The citations below are real but come from different patients in the studies. They are here represented by the hypothetical person "Amy."

The case-type is chosen because anorexia nervosa is commonly used as a paradigm example of the complexities involved with inauthenticity judgments. The aim when designing the case-token has been to reflect the difficulties of authenticity-related moral problems that are sometimes a reality in healthcare settings. Although the case is purposefully designed for a specific theoretical cause, it is realistic. The realism is central for the present purposes, which is why the case is not designed through mere speculation but is based on empirical studies. To the best of my understanding, "Amy" is a truthful representation of real persons who have been diagnosed with anorexia nervosa.

The hypothetical Amy

Amy is 25 years old. She has been diagnosed with anorexia nervosa but is now recovered. Two years ago, Amy had a body mass index (BMI)

of 17. She then visited a psychiatrist regularly but did not want any physical treatment or medication related to her anorexia. The interview with Amy includes questions which are pertinent to her authenticity and identity, to her decisions and decision-making capacity, and to her values and self-appreciation.

Here, she describes her disorder as separate from her real self (Hope et al. 2011, p. 22):

> (1) It IS like another voice, it is like another, it's almost like having two bits of you that are you all the time. The bit of you that is really scared of food and everything that means and the rest of you that wants to be able to get on without it. I just feel like there's two voices in my head sometimes.

> (2) So I didn't really want treatment, but then there's this little voice deep down inside, which is kind of the complex part, that's saying "you know you do want treatment really," but then there's this kind of overriding big THING which is just like "no, you're FAT" (laughs), "you don't need to put on weight!"

Here, Amy describes how her disorder influenced her personality (ibid, p. 23):

> (3) I feel like it's [the anorexia nervosa] made me a meaner person than I was before, [...] it's really weird because at some times I can be, like, the most selfless person [...] and other times I can be completely selfish.

And, here she describes how her desires are conflicting (ibid, p. 24):

> (4) But at the moment it's really hard, I want to eat the normal amounts, but it's really hard because at the moment, if I did eat the normal amounts I know that I wouldn't feel happy about it. But I want to be able to.

Here, she describes her anorexia as part of her identity (ibid, p. 25):

> (5) Once you've taken that [the anorexia nervosa] away, you've taken away part of my identity, so I'm bound to feel a bit lost. [...] It's like you're trying to take away the something that is a huge part of my life, [...] and if that goes what am I left with?

Here, Amy describes a difficulty to apply a factual belief to her own situation (Tan et al. 2006 p. 271):

> (6) There's part of me that didn't believe it [risk of death], but then I did feel very ill. [...] Because I didn't get to an incredibly, incredibly low weight, I wasn't in hospital, so in which case, I thought, "ok, maybe half a stone down the line that would be very, very true but at the moment I don't think it's going to happen." But also at that point it was a very focused and not very happy life so to be honest I also didn't care.[7]

This is how Amy answers a question on how important her weight and body size is to her (ibid, p. 274):

> (7) I suppose if I were answering the question for anyone else I would probably say it was of no importance, because all my friends are of different sizes and it doesn't make any difference, but just for me it's different, I feel like I suppose because I got so caught up in it that it is really important, but I don't know why, but it is; I feel really guilty of myself, putting weight on it puts on it makes me feel really different.

To summarize, Amy reports of conflicting identities and desires. She explains how the disorder had an influence on her personality, her capability to appreciate the nature of her situation, and on her values and self-appreciation. Now, the analytical task is to make reliable judgments of inauthenticity. When Amy was in this condition she declined medical treatment. Was that decision based on inauthentic desires?

[7] 1 stone = 14 lb = 6.4 kg.

Analyzing Amy

Amy's case is complex. She is an adult and under normal circumstances thus both legally and morally entitled to make her own healthcare decisions. Yet, it is clear that there are serious autonomy-related problems involved with her case. It is of interest whether Amy's healthcare decisions should have been overridden in concern for her well-being. It should be recognized that the outcome of Amy's case is already known; she is now recovered. This may influence our intuitions. But, the analytical task concerns decisions that Amy made while she was ill, so the outcome must be set aside.

The analysis must begin with answering to what kind of person Amy is and what kind of desires it is that it subject to critical scrutiny. With the theory that is presently being spelled out, it is crucial that Amy's medical condition was such that it could have influenced her decision-making so that she hurt herself or others, and that her healthcare decision was high-stake accordingly. This was true in Amy's case, which is reflected in the diagnostic criteria of anorexia nervosa (American Psychiatric Association 2013) and in citations 6 and 7.

To proceed, the substantive analysis of the desires under scrutiny concerns whether they are due to causal factors that are alien to Amy and whether they deviate from her practical identity.

Clearly, anorexia nervosa is one major causal factor. It is a disorder and as such it is alien to how Amy is otherwise construed, taking both physical and mental dispositions into consideration. It is at least partly because of alien causes that Amy declined medical treatment. To some extent, the causal history of Amy's desires indicate that they are inauthentic, i.e., that she would disapprove of having them upon critical and informed self-reflection.

It is less clear that the underlying desires are incoherent with Amy's practical identity, that is, how her identity has developed over time and was sustained at the time that she made her healthcare decisions. In citations 1 and 2, she reports of a duality of her personhood, but the descriptions are vague and do not support anything conclusive regarding her practical identity. But, in citation 3 she expresses the view

that the disorder had influenced her personality in a way that she did not appreciate. Thus, when she was ill, she held at least some desires that deviated from her practical identity. And, in citation 4 Amy says that there is a "normal" amount of food, here interpreted as normal *to her*, and that it was hard for her to eat so much. This report implies not only that she held desires that were internally conflicting, but also that they conflicted with who she really was. In light of these observations, Amy's desire to decline medical treatment appear to be deviating and, thus, possibly inauthentic. However, the analysis is not complete yet.

In citation 5, Amy explicitly states that anorexia nervosa is part of her, meaning that the disorder and its influences are not deviating from her practical identity. And, in citation 7, she reports that her weight and body size is very important to her. These values may be due to her anorexia nervosa, but it may also be the case that her anorexia nervosa is due to Amy having these values. In light of these observations, Amy's desire to decline medical treatment instead appear to be coherent with her practical identity. Or, it is at least not obvious that the desire is incoherent.

Thus, Amy's desires are partly conflicting with how her identity has developed over time and was sustained at the time that she made her healthcare decisions. To some limited extent, this conflict indicates that the desires are inauthentic, i.e., that Amy would disapprove of having them upon critical and informed self-reflection. However, Amy's claims that the disorder is part of her and that her weight and body size is very important to her indicate the opposite, i.e., that the underlying desires of Amy's healthcare decision are in fact authentic. These contradictory indicators should give rise to further investigation and follow-up questions. This is unfortunately impossible in the present case, as there is no more information available. Therefore, although there is some limited evidence supporting the judgment that Amy's desires are inauthentic, the factors from deviation which are available for analysis are inconclusive.

In conclusion, both factors of alien causes and of deviation are present in the case of Amy. However, because of the epistemic uncertainty involved, which are due to contradictory evidence, it is only justified to some limited extent to judge that the desires underlying

her decision to decline medical treatment are inauthentic and, thus, non-autonomous to some degree.[8]

Evaluating the test

The first thing to be noted is that the analysis is fruitful even though the case is complex, vague, and contains little detailed information. The central authenticity-related moral problems are clearly articulated with solid theoretical support. Both factors that indicate inauthenticity and factors that indicate the opposite are explicated in detail, which enables critical scrutiny. Also, the results appear to be generally reason-giving in a framework of reflective equilibrium. That is, to some limited extent, the analysis supports the judgment that Amy's decision was non-autonomous for authenticity-related reasons. The theory is successful in these respects, mainly because it provides a conceptual framework that enables detailed analysis.

Furthermore, it seems reasonable to assume that the theory could be even more fruitful in future analyses of similar cases, as it provides theoretical support for focusing on a certain kind of behavior in patients and for asking them certain kinds of questions. It provides a reliable framework for analyses of personal autonomy in terms of authenticity.

Concluding remarks

In this article, the Frankfurt–Dworkean tradition of thinking about authenticity has been merged with Beauchamp and Childress's non-ideal account of autonomy. The result is a non-ideal authenticity-based account of autonomy that, when applied, seems to be fruitful in analyses of the degree of autonomy of patients' decision-making in healthcare. Thereby, the theory succeeds in providing reliable and practical action-guidance in a matter which has concerned theorists and practitioners for some time.

[8] Citation 6 reflects that Amy's decision may have been non-autonomous to some extent also in the sense that she had limited decision-making capabilities, but it is beyond the present purposes to elaborate on this observation.

Because it provides reliable action-guidance, the theory may be foundational for paternalist considerations of coercive care. However, such considerations must include normative support concerning the sufficient degree of epistemic certainty of inauthenticity and the justified proportionality of the intervention, among other things. Thus, although the present theory may be foundational for paternalist interventions in the name of authenticity, it does not by itself provide sufficient moral support for coercive care.

References

Ahlin, J. (2018a). The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent. *Medicine, Health Care and Philosophy* 21(1), 43–50.

—. (2018b). What Justifies Judgments of Inauthenticity? *HealthCare Ethics Committee Forum* 30(4), 361–377.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5 ed.). Washington, DC: APA.

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of Biomedical Ethics* (7 ed.). Oxford University Press.

Beauchamp, T. L., & Rauprich, O. (2016). Principlism. In H. ten Have (Ed.), *Encyclopedia of Global Bioethics* (pp. 2282–2293). Springer.

Buchanan, A. E., & Brock, D. W. (1990). *Deciding for Others: The Ethics of Surrogate Decision Making*. Cambridge University Press.

Burns, J. M., & Swerdlow, R. H. (2003). Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign. *Archives of Neurology* 60(3), 437–440.

Christman, J. (2009). *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge University Press.

DeGrazia, D. (2005). *Human Identity and Bioethics*. Cambridge University Press.

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1), 5–20.

Hope, P. T., Tan, D. J. O. A., Stewart, D. A., & Fitzpatrick, P. R. (2011). Anorexia Nervosa and the Language of Authenticity. *Hastings Center Report* 41(6), 19–29.

Juth, N. (2005). *Genetic Information: Values and Rights. The Morality of Presymptomatic Genetic Testing.* Acta Universitatis Gothoburgensis.

Lester, R. J. (2009). Brokering Authenticity Borderline Personality Disorder and the Ethics of Care in an American Eating Disorder Clinic. *Current Anthropology* 50(3), 281–302.

Noggle, R. (2005). Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (pp. 87–108). Cambridge University Press.

Sacks, O. (1985). *The Man Who Mistook His Wife For a Hat: And Other Clinical Tales.* Touchstone Books.

Sjöstrand, M., & Juth, N. (2014). Authenticity and Psychiatric Disorder: Does Autonomy of Personal Preferences Matter? *Medicine, Health Care and Philosophy* 17(1), 115–122.

Swanson, K. R. G. (2017). *Non-Ideal Autonomy in Relationships of Care* (doctoral thesis). University of Minnesota.

Swindell, J. S. (2009). Two Types of Autonomy. *The American Journal of Bioethics* 9(1), 52–53.

Tan, D. J. O. A., Hope, P. T., Stewart, D. A., & Fitzpatrick, P. R. (2006). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. *Philosophy, Psychiatry, & Psychology: PPP* 13(4), 267–282.

Taylor, J. S. (2005). Introduction. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy* (pp. 1–29). Cambridge University Press.

Valentini, L. (2012). Ideal vs. Non-ideal Theory: A Conceptual Map. *Philosophy Compass* 7(9), 654–664.

White, L. (2018). The Need for Authenticity-Based Autonomy in Medical Ethics. *HealthCare Ethics Committee Forum* 30(3), 191–209.

 The article is originally published as: Ahlin Marceta, J. (2018). A Non-Ideal Authenticity-Based Conceptualization of Personal Autonomy. *Medicine, Health Care and Philosophy* DOI:doi.org/10.1007/s11019-018-9879-1. **The following changes have been made in the present version:** (1) The list of references has been updated according to the reference standards of this thesis. (2) The words "of an earlier version of this article" have been removed from footnote 6 on p. 102. (3) The letter s has been added to the word "concept" on p. 95. **Acknowledgements:** This study was partly funded by the Swedish Research Council and the Swedish Research Council for Health, Working Life and Welfare (Grant number 2014–4024). I am also grateful to SciencesPo and the CEVIPOF institute in Paris, France, for hosting me while I finished writing this paper.

Nine Cases of Possible Inauthenticity in Biomedical Contexts and What They Require from Bioethicists

Abstract: Respect for autonomy is a main moral principle in bioethics. It is sometimes argued that authenticity, i.e., being "real," "genuine," "true to oneself," or similar, is crucial to a person's autonomy. This article collects nine cases in which the notion of authenticity has been or could be invoked in biomedical contexts. One recently developed theory aiming to provide normative guidance with regard to authenticity-related problems is applied when it is possible, while it is explained in detail why the theory is inept or impractical in the remaining cases. The article thus provides an overview of authenticity-related problems which may be helpful for autonomy theorists. Furthermore, it is argued that there is no universal problem of authenticity, but many problems, and that they may require various particular solutions rather than one universal solution. Among other things, it is suggested that bioethicists should explore non-ideal methodological approaches to authenticity-related problems to provide action-guidance with regard to them.

Keywords: Authenticity, autonomy, healthcare, bioethics

Introduction

"I wasn't really bothered about dying, as long as I died thin." The citation is an excerpt from an interview conducted with a person who talks about her anorexia nervosa (Tan et al. 2006, p. 274). The person reports that being thin was more important to her than being alive. Is her wish authentic? Is it really *hers*, in a substantive sense? The question has engaged bioethicists and medical practitioners, partly because the answer to it may also be important to another question, namely whether the person's healthcare decisions should have been respected.

This article collects nine real and hypothetical cases in which bioethicists and medical practitioners have found the notion of authenticity morally relevant in judgments of patients' decision-making, or could reasonably be expected to find the notion relevant accordingly. It is argued that there are many different authenticity-related problems that require different approaches, and that no theory of authenticity that is present in the contemporary bioethical literature is capable of providing universal guidance with regards to all of those problems. The article begins with an introduction to authenticity theory and its role in biomedicine. The section also introduces a recently developed theory aiming to provide normative guidance with regard to authenticity-related problems, namely Ahlin Marceta (2018). In the subsequent section, nine authenticity-related problems are accounted for, including comments about what is required from authenticity theorists to solve them. A brief final section concludes.

Authenticity in biomedicine

The moral concern

To be autonomous is to be self-governed (Christman 2015). Respect for autonomy is one of the main moral principles in contemporary bioethics (cf. Beauchamp and Childress 2013). In concern for patients' autonomy, bioethicists invoke concepts such as decision-making capacity (Grisso et al. 1997) and voluntariness (Nelson et al. 2011). That is, if a patient is not capable of making healthcare decisions, or if she is not making healthcare decisions which are independent from undue influences such as social or economic pressures, this has a negative effect on the degree of autonomy of her healthcare decisions.

During the 20th century, informed consent practices have been incorporated in healthcare in large parts of the Western world with the aim of respecting and promoting patient autonomy (Jonsen 2000; Faden and Beauchamp 1986). In recent years, various bioethicists have raised the possibility of incorporating authenticity in autonomy-based practices in healthcare (Ahlin Marceta 2018; Sjöstrand and Juth 2014;

White 2018). It is not entirely clear how the notion of authenticity should be conceptualized, although the term is usually understood to mean "genuine," "real," "true to oneself," or similar.

The bioethicists' concern has been that healthcare decisions must be authentic to be fully autonomous. Among the problems associated with this concern is that authenticity is difficult to detect in others (Ahlin 2018a; Sjöstrand and Juth 2014). More specifically, it is difficult to justify the judgment that someone else's person or decision is inauthentic (Ahlin 2018b). Furthermore, although there may be one true conception of authenticity, it is likely that real authenticity-related problems require different kinds of solutions. Thus, bioethics may not need one universal theory of authenticity but various theories that explain and solve different authenticity-related problems.

The present article supports that view. Here, nine authenticity-related problems are explicated as they have been (or could reasonably be) treated by bioethicists and medical practitioners. It is argued that there is no universal problem of authenticity, but many problems, that they must be framed differently and, thus, solved differently. It is concluded that bioethicists have reason to engage in authenticity-related problems with aims and approaches that are specific for the particular problem at hand, and explore the possibility of taking a new non-ideal methodological grip on them.

Theories explaining authenticity

The perhaps most prominent tradition of thinking about authenticity has its roots in a series of books and articles from the 1970's and 1980's, of which Frankfurt (1971) and Dworkin (1988) may be the most noteworthy. In this tradition an act, decision, or desire is authentic if the agent endorses it on a higher level of reflection. For illustration, consider a drug addict who has two conflicting wishes on two different levels of desire. On one desire-level, she wants to shoot heroin. On a higher desire-level, she wants to lead a long and healthy life. The desires are conflicting, and because of that conflict the desire on the lower level is deemed inauthentic. One criticism of so-called split-level theories of

authenticity is that desires on the higher level must also be endorsed on a yet higher level to be authentic, and desires on that level must also be endorsed on a yet higher level, and so on in an infinite regress (cf. Taylor 2005). If the critics are right, there is something inherently problematic with the kind of authenticity theories which have gained most attention from philosophers and bioethicists in recent decades.

Other theories of authenticity include, for instance, such that put weight on the causal history of desires and such that focus on the coherence of full desire-sets. Elster's theory is one example of the former. In it, desires are inauthentic if they are "shaped by irrelevant causal factors, by a blind psychic causality operating "behind the back' of the person" (Elster 1983, p. 16). In this line of thought authentic desires have a certain kind of origin, most often in some cognitive processes of the desire-holder (Ahlin 2018, p. 46). One example of a coherence-oriented theory is found in Miller, who writes that authentic actions are "consistent with the person's attitudes, values, dispositions, and life plans" (Miller 1981, p. 24). In this line of thought actions, decisions, or desires are instead authentic if they are coherent with the desire-holder's full set of desires (Ahlin 2018, pp. 46–7).

However, these theories are all oriented around decision-making or acting. Bauer (2017) offers an alternative approach, namely the focus on what it is to be an authentic person. The ideal of being an authentic person, in Bauer's proposal, is a combination of the ideal of expressing and unfolding one's individual personality and the ideal of being an autonomous person who is morally responsible (p. 579). In more elaborate terms, the ideal is comprised of (1) aspects of being authentic by being a self with distinctive characteristics of an individual personality. These aspects include the free unfolding of one's individual personality, expression of oneself in acting and living, and being true to one's own convictions, beliefs, ideals, life-plans, and projects (ibid). Furthermore, the ideal is comprised of (2) aspects of being authentic by being "a person" in terms of an autonomous (moral) agent. These aspects include giving reasons and taking moral responsibility for one's actions, being a reflective "self-evaluator," and being a trustworthy partner of social interaction (ibid).

One other alternative is to focus instead on what it is to lead an authentic life (cf. Taylor 1991). However, these alternative approaches have not gained as much attention from bioethicists as the desire-oriented approach, perhaps because bioethicists' main focus is on autonomous decision-making. It will be shown below that some authenticity-related problems are difficult to phrase in terms of decision-making, while others are difficult to *not* phrase in such terms.

A recently developed normative theory of (in-)authenticity judgments

In Ahlin Marceta (2018), I suggest a desire-oriented theory aiming to provide guidance in practical authenticity-related problems. According to the theory, the relevant problem is to justify judgments that someone's healthcare decision builds from inauthentic desires.[1] For reasons of justification, the theory is delimited to concern "persons whose medical condition may influence their decision-making so that they hurt themselves or others." For such persons, and their possibly harmful healthcare decisions, "it is justified to judge that an underlying desire is inauthentic to the extent that it is due to causal factors that are alien to the person and to the extent that it deviates from the person's practical identity."

In this theory, two factors must be present for a judgment of inauthenticity to be justified:

> **The factor of deviation** It is a factor indicating inauthenticity that the desire under scrutiny does not cohere with how the desire-holder's identity has developed over time and is presently being sustained.
>
> **The factor of alien causes** It is a factor indicating inauthenticity that the desire under scrutiny is due to causes that are not normal to how the desire-holder is otherwise construed, taking both physical and mental dispositions into consideration.

Both factors are expressed in degrees rather than in necessary and sufficient conditions, and are sensitive to judgment. It is, for instance, not

[1] The article is published as pre-print and lacks page numbers.

stated *a priori* what it means for a cause to be "not normal" to how the desire-holder is otherwise construed. The theory requires practical and context-sensitive deliberation in particular cases.

Its application is a two-step process. First, it must be determined whether the person whose healthcare decisions are evaluated suffers from a medical condition that may influence their decision-making so that they are harmful to themselves or others. Second, it must be determined whether the two factors are present, and if so, to what extent. In Ahlin Marceta (2018), the process is demonstrated on a hypothetical (but empirically grounded) case of anorexia nervosa.

In the below section, it is argued that the theory can be fruitfully applied in three of the authenticity-related cases discussed (case 1, 2, and 9), but that it is inept in the six remaining cases.

Nine authenticity-related cases

Overview

The cases are (1) inauthenticity from physical causes, (2) inauthenticity from psychological causes, (3) unstable desire-sets, (4) lack of desires, (5) medically induced authenticity, (6) inauthentic recovery, (7) indoctrinated desires, (8) false selves, and (9) unexplained surprising desires. Cases 1 through 5 build on actual cases while cases 6 through 9 are hypothetical.

Case 1: Inauthenticity from physical causes

In a case study, Burns and Swerdlow (2003) report of an otherwise normal 40-year old man who suddenly developed a sexual interest in children. The man had no previous pedophilic symptoms, and did not want to have them either; among other things, he underwent a 12-step program for sexual addiction to be able to lead a normal life. Upon medical examination, it was found that the man's sexual desires were due to a brain tumor. He had developed a right orbitofrontal tumor which affected him cognitively and behaviorally. When the tumor was removed, the pedophilic symptoms disappeared. When the symptoms later returned, it was found that that the tumor had done so too. Thus,

there is a clear and unambiguous causal relationship between the man's brain tumor and his sexual desires. There seems to be authenticity-related problems connected to the case.

One way to phrase one such problem is that the man's sexual desires do not seem to be authentic. Another phrasing is that the man does not seem to be an authentic pedophile. It is not immediately clear whether the two phrasings are substantially different. A theory of authenticity that is oriented around decision-making would support the former phrasing, while a theory that is oriented around personhood would support the latter.

If the problem is understood as concerning decision-making, the theory from Ahlin Marceta (2018) can be fruitfully applied to it. First, the man's medical condition could have influenced his decision-making negatively in the sense described by the theory. This is obvious from the case description. Second, both the factor of deviation and the factor of alien causes are present. The generic case description above does not state to what extent they are present. However, that is not required for the present purpose, which is to consider whether the theory can be fruitfully applied to cases of authenticity from physical causes. It should be reasonably clear from this brief discussion that the theory is applicable in such cases, although its full potential can only be realized in more detailed particular instances.

Case 2: Inauthenticity from psychological causes

Anorexia nervosa is usually treated as a psychiatric disorder. However, it should be noted that patients suffering from it can be fully competent to make healthcare decisions. Many can understand information relevant to their condition and the recommended treatment, reason about the potential risks and benefits of their choices, appreciate the nature of their situation and the consequences of their choices, and so on. Yet, they assess their own bodies, i.e., mainly their weight and physical appearance, unreasonably. Consider this excerpt from an interview conducted with an anorexia nervosa patient. It is representative also of other interviews in the same article (Tan et al. 2006, p. 274):

> *Interviewer*: What is the importance of your weight and body size to you? "I just want to be thin." *Interviewer*: How important is that to you? "Very." *Interviewer*: Why? "It just is, it's all I want."

Thus, some anorexia nervosa patients have wishes that appear to be defective in some way, not as a matter of incompetence but of values. It is a problem to determine on what grounds these wishes are defective, and one suggestion is that it is because they are inauthentic.

Many would make the intuitively valid claim that the patient has inauthentic wishes because she has anorexia nervosa. However, inauthenticity is not listed among the diagnostic criteria for the disorder (see, e.g., American Psychiatric Association 2013). Therefore, although the patient's wishes may be inauthentic, it is not because she has anorexia nervosa but for some reason external to the disorder. The intuitively valid claim that the patient's wishes are inauthentic because she is anorectic is thus not empirically or conceptually valid. It could reasonably be argued that inauthenticity should be among the diagnostic criteria of anorexia nervosa, although it then remains to explain precisely what it is for something or someone to be inauthentic.

It may also be argued that our intuitions are misguided or misinterpreted in this case. They are not intuitions about the possible inauthenticity of the patient's wishes, but about the patient's welfare. That is, the intuition is in fact that the patient's wishes are defective because it is not good to have them. Obviously, this can be true for some readers. Yet, various clinicians and bioethicists, such as, e.g., Hope et al. (2011), Sjöstrand and Juth (2014), and Tan et al. (2006), have expressed and analyzed the possible problem of anorexia nervosa patients' wishes in terms of authenticity. Their analyses do not appear to rest on misguided or misinterpreted intuitions, but on the considered view that there is some authenticity-related problem with such wishes.

The target case in Ahlin Marceta (2018) is precisely a case of anorexia nervosa, and I will not repeat the analysis here. It should be sufficient to declare that the theory is (arguably) fruitful also in cases where there appears to be problems connected to wishes that are intertwined with the diagnostic criteria of some disorder.

Case 3: Unstable desire-sets

Among other things, patients suffering from borderline personality disorder (BPD) are characterized by unstable "selves," which has prompted ethicists to consider the ethics of caring for BPD patients in terms of authenticity (Lester 2009). A BPD patient could, for instance, display sudden and dramatic shifts in goals, values, vocational aspirations, types of friends, and so on (ibid, p. 284). In extreme situations, BPD patients can make a series of mutually incompatible healthcare decisions resting on unstable desires. For instance, a BPD patient may request forced medication, as only that enables her to go through psychotherapy, and minutes later refuse medication, as one of its side effects is that it clouds her thinking. Healthcare personnel cannot adhere to both wishes.

The main authenticity-related problem in this case appears to be that BPD patients have too unstable desire-sets. Surely, a normal person could have authentic but conflicting wishes in subjects of minor importance, such as an authentic wish to eat ice cream and an authentic wish to not eat sugar. Also, normal persons could reasonably be authentically indecisive, at least to some extent. But BPD patients appear to be unstable in a way that calls for judgments of inauthenticity. That is, there is a seriousness to their symptoms that makes it reasonable to assess their personality, or their decisions, in terms of authenticity. However, it remains for theorists to explain precisely why and how their instability is an authenticity-related problem, if at all.

The theory in Ahlin Marceta (2018) does not appear to be capable of treating the main moral problem in this case. The theory could be applied to particular decisions made by BPD patients, although the problem is not the decisions *per se* but that they rest on unstable desire-sets. Therefore, provided that this instability is an authenticity-related problem, some other theory than Ahlin Marceta (2018) must be developed to treat it.

Case 4: Lack of desires

The late stages of schizophrenia may include "negative" symptoms such as underactivity, blunting of affect, passivity, and lack of initiative (American Psychiatric Association 2013). Schizophrenics in this stage can

sometimes lead reasonably normal lives, while being completely indifferent to anything that happens to them and how their lives go. It does not matter to them whether they are healthy, live in a comfortable home, or have meaningful relationships with others. They can be described as living without any wishes.[2]

The question can be raised whether this condition is authentic, i.e., whether a person can authentically lack wishes. In some cases a state of mind which is free of wishes is desirable, such as when it is the wanted result from deliberate meditation. Buddhists, mindfulness practitioners, and others, seek to not have any desires. However, it is different to be in that condition due to some medical disorder. Thus, it is a problem for authenticity theorists to clarify whether it is possible to authentically lack wishes, where this lack is due to some disorder, and if so also why.

Furthermore, when these questions have been resolved, a theory must be developed that can be applied to reliably determine whether a desire-free condition or state of mind is inauthentic. As the problem here is not to determine whether any particular decision rests on inauthentic desires, the theory from Ahlin Marceta (2018) cannot be applied for guidance.

Case 5: Medically induced authenticity

In the first chapter of his book *Listening to Prozac* (1993), Kramer reports of Tess, a patient whose personal story is extraordinary. Among many other things, Tess was a victim of child abuse. She suffered from depression and had suicidal thoughts (p. 3). After various failed attempts at medication and therapy Kramer prescribed Prozac, which at the time had recently been released by the U.S. Food and Drug Administration. Soon thereafter, Tess showed a remarkable change. Her work became more satisfying, her social relationships changed to the better, and she was "astonished at the sensation of being free from depression" (p. 7). After nine months, Tess went off medication and continued doing well. About eight months after that, she told Kramer that she was slipping. She said, "I'm not myself" (p. 10). Thus, Prozac made Tess authentic (per self-report).

[2] I adopt this characterization from dialogues with psychiatrists.

The case draws out a conflict of intuitions. On the one hand, it is intuitive to hold that Tess's self-reports of authenticity are real simply because they are self-reported. On the other, it is counterintuitive to hold that she is authentic, as it is known that her condition is induced by medication. Ahlin Marceta (2018) is not helpful here, as the theory is not intended to answer to the questions presently being asked. There is thus reason for authenticity theorists to organize and explain these conflicting intuitions in new theoretical work.

One possible explanation of the case is that Prozac helped Tess to "find" the authentic self that she was before she was abused as a child (provided that the abuse caused the inauthenticity). However, this explanation is more complex than what first appears.

In one sense, Tess pre-abuse is not the same person as Tess post-abuse, because the former is a child and the latter is an adult. If Prozac helped Tess to "find" the authentic self that she was before she was abused, its effect is very specific; Prozac did not affect features of Tess's personhood that are connected to her being an adult, but only features that are connected to some core of authenticity in her as a person. Thus, the explanation assumes that Prozac, in this case, had an extremely accurate medical effect. Furthermore, the explanation rests on the assumption that authenticity concerns something that does not change over time, namely some personhood-related entity which remains the same in both Tess pre-abuse and in Tess post-abuse. Thereby, it commits to theories of personhood, philosophy of mind, and possibly also phenomenology, according to which a person is something intertemporally fixed. These theories are not obviously true. Thus, the explanation is simple and attractive at first glance, but upon closer examination it becomes clear that it carries a large theoretical load which makes it very complex.

One other possible explanation is that Tess confuses who she is with who she wants to be. She wants to be the person that Prozac helps her to be, and therefore she states that this person is who she really is. This explanation is also more complex than what first appears. If it is correct, normally informed and competent persons can be mistaken about who they really are, in terms of authenticity. The explanation may disqualify

theories of authenticity that are oriented around self-assessment, and which have otherwise been prominent in authenticity theorizing since Frankfurt (1971) and Dworkin (1988).

In conclusion, intuitively reasonable explanations of the case with Tess are theory-dependent and complex upon closer examination. It remains for authenticity theorists to treat cases of medically induced authenticity in greater detail.

Case 6: Inauthentic recovery

Some disorders can be treated with either medicine or psychotherapy (or both). It can be argued that, for reasons of authenticity, psychotherapy is a better option than medicine. This line of thought has been explored by, e.g., Kass (2003, pp. 22–3):

> In most of our ordinary efforts at self-improvement, either by practice or training or study, we sense the relation between our doings and the resulting improvement, between the means used and the end sought. There is an experiential and intelligible connection between means and ends; we can see how confronting fearful things might eventually enable us to cope with our fears. We can see how curbing our appetites produces self-command. [...] In contrast, biomedical interventions act directly on the human body and mind to bring about their effects on a subject who is not merely passive but who plays no role at all. [...] The relations between the knowing subject and his activities, and between his activities and their fulfillments and pleasures, are disrupted.

It is one argument that psychotherapy is better than medicine because of some positive secondary effects, such as a strengthened self-esteem or longer lasting medical result. I am not concerned with that here. But, it can also be argued that psychotherapy is better than medicine because of some authenticity-related reason. That is, the opinion is feasible that authentic recovery from disorder is better than inauthentic recovery. But, the opinion rests on the idea that there is such a thing as inauthentic

recovery, and it is not immediately clear that there is theoretical support for this idea beyond mere intuition.

This is different from questions of whether someone's decision between treatment and therapy is authentic. The problem for theorists, if it is a problem at all, is to make a clear and unambiguous distinction between authentic and inauthentic recovery processes.[3] Obviously, Ahlin Marceta (2018) is not useful here.

Case 7: Indoctrinated desires

Consider this thought example (Taylor 2005, p. 11):

> [Imagine] a child at time t whose mother wished him to learn to play the piano and who beat him if he did not practice. As time passes and the child grows more proficient at playing, he discovers (at time t1) that his mother's belief that piano playing suited him was right, and he comes to love playing – even though he still repudiates the means by which his mother brought him to this position.

The thought example is intended to bring out a conflict of intuitions; intuitively, the man's love for playing the piano is formed in the wrong way and is therefore inauthentic, but the man endorses his own love for playing the piano upon informed and critical self-reflection and therefore it is intuitive to hold that it is authentic.

Different authenticity theories explain such cases of manipulation or indoctrination differently. Theories that emphasize the causal history of desires, such as, e.g., Elster's (1983), would determine that the child's love for playing the piano is inauthentic. Theories that focus on self-affirmation, such as, e.g., Frankfurt's (1971) and Dworkin's (1988), would instead determine that the child's love for playing the piano is authentic.

One more straightforward example of indoctrination is discussed by Robert Noggle (2005, p. 102):

[3] See also Svenaeus (2009), who has previously argued that there is no ethically relevant difference between psychopharmacological and psychological self-change.

> *Edgar the Evil* is a son of a crime boss who rears him to follow in his footsteps. Using standard child-rearing techniques, he encourages Edgar's more selfish and violent impulses and discourages empathy and compassion. As Edgar reaches adulthood, he is quite thoroughly evil.

The commonly shared intuition is that Edgar is not authentically evil. Edgar the Evil is analogous to people who, for instance, grow up in religious sects or live under oppressive patriarchic circumstances. Sometimes such people make dubious healthcare decisions that indicate inauthenticity. For instance, many bioethicists today agree that the wishes of a Jehovah's Witness who refuses blood transfusion should be respected for anti-paternalist reasons. Further analysis may be feasible concerning their possible inauthenticity; perhaps there are similar cases in which reliable indicators of inauthenticity provide sufficient grounds for paternalist interventions.

It remains for authenticity theorists to organize and explain the various conflicting intuitions in cases of manipulation or indoctrination, and to provide clear and unambiguous action-guidance with regard to them. The theory in Ahlin Marceta (2018) is partially guiding here, but it does not answer the relevant questions. Presumably, neither manipulation nor indoctrination are medical conditions. Therefore, manipulated or indoctrinated patients are not the kind of persons that, according to Ahlin Marceta (2018), are justifiably targeted by inauthenticity judgments. However, this normative guidance is not satisfying. It side-steps the relevant moral problem, namely the possible inauthenticity of decisions that are due to manipulation and indoctrination, rather than solves it.

Case 8: False selves

Winnicott (2007) introduced a thought example called the "False Self" which has been used as a paradigm model of inauthentic behavior (see, e.g., Velleman 2002, pp. 97–8). In the example, we are to picture a person who "laughs at what he thinks he is supposed to find amusing, shows concern for what he thinks he is supposed to care about, and in general conforms himself to the demands and expectations of others" (Velleman

2002, p. 97). He fails to be motivated "from within his true self" and is therefore inauthentic (ibid). The lesson we are supposed to learn is that conformity, in some sense, negates authenticity. However, it is not obvious that the example is successful in showing that. Taylor comments on the False Self person that, "while his laughter might not be authentic in the sense of its expressing genuine amusement, it would be authentic in the sense of being representative of this person's other-directedness. It would be authentically inauthentic" (Taylor 2009, p. 32). In other words, the False Self person might be an authentically other-directed person.

Taylor does have a point, although there is something distressing about his remark. The False Self example draws attention to the intuition that there is something inauthentic about people who conform to what they believe to be others' wishes rather than to formulate and follow their own. But the example is too strong. Humans are socially embedded beings; everyone conforms to others' expectations to some extent, at least during periods of our lives. In many cases, we tend to think that people who fail to conform to others' expectations lack social skills. We even hope that our children learn the social balance between following one's own desires and conforming to others'. Thus, it is difficult to draw the straight and unambiguous line between "self-motivation" and "else-motivation" that the False Self is intended to illuminate. However, the thought of a person who is "authentically inauthentic," as Taylor suggests, is as distressing as being completely insensitive to the expectations of others. In reality, the normal case is likely that authentic people are somewhere in between fully self-motivated and fully else-motivated.

There is disagreement among authenticity theorists regarding problems that are connected to the tension between social influences and the self. It is possible that the main merit of the False Self example and Taylor's comments is that they illuminate one problem associated with constructing a hypothetical ideal of authenticity; perhaps any ideal model of authenticity would be torn apart by the forces in the dialectics above. No person can be either authentically fully self-motivated nor authentically fully else-motivated, and therefore any ideal that is oriented around either extreme is inherently flawed. Instead, it may be argued, a theory

of authenticity should be non-ideally constructed, and account for the tension between social influences and the self already from the outset.

The theory in Ahlin Marceta (2018) is non-ideal in this sense. However, as in the above, the problem presently described is not of the kind that Ahlin Marceta (2018) is intended to solve.

Case 9: Unexplained surprising desires

Consider the hypothetical case of Anna, "a young and promising professional ballet dancer" (Ahlin 2018, p. 44). Anna loves her work, has moved across the nation to attend the best ballet schools, set aside personal relationships that conflicted with her career, and is known by those who are close to her to love dancing more than anything else. In the case, Anna has suffered a serious leg injury and must undergo a minor surgery to avoid implications that will in time necessitate an amputation. Anna is competent to make healthcare decisions and is fully informed about the consequences of her decisions, yet she refuses to undergo surgery. Her treating clinician reflects upon the case and believes that Anna's decision rests on inauthentic desires.

The case is intended to illustrate that it is often surprises that bring attention to the notion of authenticity; as long as people make decisions that are not unexpected, we do not consider them in terms of authenticity. But, with support from Ahlin Marceta (2018), the case also shows that decisions are not inauthentic merely because they are surprising, not even if the decisions are surprising to the extent that they conflict with everything that is known about the decision-maker. Judgments of inauthenticity require a real and elaborate explanation. In the case of Anna, the causal history of her desires are unknown and therefore the requirement to meet the factor of alien causes is not fulfilled. Thus, the theory in Ahlin Marceta (2018) provides guidance here.

Lessons to be learned

Authenticity issues relate to a number of different problems. In some of the cases above, the main problem of authenticity is related to decision-

making. In others, the problem rather concerns personhood or being in some condition. Therefore, there is likely no universal solution to authenticity-related problems, but various particular solutions.

As mentioned briefly in the discussion of case 8, it is possible that bioethicists should further consider a non-ideal methodological approach to authenticity-related problems. Most (or all) theories of authenticity are comprised of some hypothetical ideal of authenticity, in the sense that they are constructed of propositions such as "X authentic if and only if Y." Then, the theories suggest that practitioners should scrutinize X's (i.e., desires, lives, persons, etc.) and observe whether and to what extent they have or are Y. It may instead be fruitful to follow Ahlin Marceta (2018) and adopt a non-ideal approach. Such approaches, which are sometimes also described as "realist," "problem-oriented," or "bottom-up," may start from the case at hand rather than from some hypothetical model of authenticity and attempt to describe what is problematic about it in particular terms. Bioethicists should at least explore the possibility of taking a new methodological grip on authenticity-related problems.

Furthermore, it may be the case that the solution to any particular authenticity-related problem must be goal-oriented, in the sense that it matters to the solution why it is interesting to solve the problem. That is, in most (or all) cases above, the main concern is related to paternalism. Therefore, the paternalist intention makes a difference to how the problems should be solved. In case 9, for instance, it is interesting to explain the possible inauthenticity of Anna because of a concern for her practical identity and way of life as a professional ballet dancer. Perhaps this concern, rather than some pre-established theory of authenticity, should be guiding in an analysis of the case. However, because the paternalist concern would then be action-guiding, it is essential that the paternalist intention is well-grounded first; the cart may only be put before the horse if this order is a moral and analytic necessity.

To summarize, this article collects nine authenticity-related cases in biomedicine. It has been argued that there is likely no universal solution to authenticity-related problems, but various particular solutions. The

theory in Ahlin Marceta (2018) provides normative guidance in cases 1, 2, and 9. Lastly, it has been proposed that bioethicists should explore alternative methodological approaches to the notion of authenticity and its applications in biomedicine. The main lessons to be learned are that there is yet a lot of analytical work to be done regarding authenticity in biomedical contexts, and that bioethicists have reason to engage in authenticity theory precisely as they have previously engaged in theorizations of concepts such as decision-making capacity and voluntariness.

References

Ahlin, J. (2018a). The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent. *Medicine, Health Care and Philosophy* 21(1), 43–50.

—. (2018b). What Justifies Judgments of Inauthenticity? *HealthCare Ethics Committee Forum* 30(4), 361–377.

Ahlin Marceta, J. (2018). A Non-Ideal Authenticity-Based Conceptualization of Personal Autonomy. *Medicine, Health Care and Philosophy*. DOI:doi.org/10.1007/s11019-018-9879-1.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5 ed.). Washington, DC: APA.

Bauer, K. (2017). To Be or Not To Be Authentic. In Defence of Authenticity as an Ethical Ideal. *Ethical Theory and Moral Practice* 20(3), 567–580.

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of Biomedical Ethics* (7 ed.). Oxford University Press.

Burns, J. M., & Swerdlow, R. H. (2003). Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign. *Archives of Neurology* 60(3), 437–440.

Christman, J. (2015). Autonomy in Moral and Political Philosophy. *Stanford Encyclopedia of Philosophy*. Retrieved (December 30, 2018) from http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge University Press.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge University Press.

Faden, R., & Beauchamp, T. L. (1986). *A History and Theory of Informed Consent*. Oxford University Press.

Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy* 68(1), 5–20.

Grisso, T., Appelbaum, P. S., & Hill-Fotouhi, C. (1997). The MacCAT-T: A Clinical Tool to Assess Patients' Capacities to Make Treatment Decisions. *Psychiatric Services* 48(11), 1415–1419.

Hope, P. T., Tan, D. J. O. A., Stewart, D. A., & Fitzpatrick, P. R. (2011). Anorexia Nervosa and the Language of Authenticity. *Hastings Center Report* 41(6), 19–29.

Jonsen, A. R. (2000). *A Short History of Medical Ethics*. Oxford University Press.

Kass, L. R. (2003). Ageless Bodies, Happy Souls: Biotechnology and the Pursuit of Perfection. *The New Atlantis* Spring(1), 9–28.

Kramer, P. D. (1993). *Listening to Prozac: A Psychiatrist Explores Antidepressant Drugs and the Remaking of the Self*. Penguin Books.

Lester, R. J. (2009). Brokering Authenticity Borderline Personality Disorder and the Ethics of Care in an American Eating Disorder Clinic. *Current Anthropology* 50(3), 281–302.

Miller, B. L. (1981). Autonomy & the Refusal of Lifesaving Treatment. *The Hastings Center Report* 11(4), 22–28.

Nelson, R. M., Beauchamp, T. L., Miller, V. A., Reynolds, W., Ittenbach, R. F., & Luce, M. F. (2011). The Concept of Voluntary Consent. *The American Journal of Bioethics* 11(8), 6–16.

Noggle, R. (2005). Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy* (pp. 87–108). Cambridge University Press.

Sjöstrand, M., & Juth, N. (2014). Authenticity and Psychiatric Disorder: Does Autonomy of Personal Preferences Matter? *Medicine, Health Care and Philosophy* 17(1), 115–122.

Svenaeus, F. (2009). The Ethics of Self-Change: Becoming Oneself By Way of Antidepressants or Psychotherapy? *Medicine, Health Care and Philosophy* 12(2), 169–178.

Tan, D. J. O. A., Hope, P. T., Stewart, D. A., & Fitzpatrick, P. R. (2006). Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values. *Philosophy, Psychiatry, & Psychology: PPP* 13(4), 267–282.

Taylor, C. (1991). *The Ethics of Authenticity*. Harvard University Press.

Taylor, J. S. (2005). Introduction. In J. S. Taylor (Ed.), *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy* (pp. 1–29). Cambridge University Press.

—. (2009). *Practical Autonomy and Bioethics*. Routledge.

Velleman, D. J. (2002). Identification and Identity. In S. Buss & L. Overton (Eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt* (pp. 91–123). The MIT Press.

White, L. (2018). The Need for Authenticity-Based Autonomy in Medical Ethics. *HealthCare Ethics Committee Forum* 30(3), 191–209.

Winnicott, D. W. (2007). Ego Distortion in Terms of True and False Self The Maturational Process and the Facilitating Environment. In D. W. Winnicott (Ed.), *Studies in the Theory of Emotional Development* (pp. 140–154). Karnac Books.

Part III

Summary / Sammanfattning

Popularized Summary in English

The purpose of this thesis is to make theoretical ideals of authenticity useful in practice. By "authenticity," I mean that someone or something is "real," "genuine," "true to oneself," or similar. The thesis includes an introduction and four articles related to authenticity in a medical context.

The problems I write about concern authenticity in decision-making, or more specifically the authenticity of desires. As other philosophers, in this context I think of "desires" as the most basic element in ordinary decision-making; a foundational attitude or directedness that influences a person's decisions. One paradigm example that is sometimes used in the philosophical literature to illustrate authenticity-related problems is a person who suffers from anorexia nervosa, and who makes healthcare decisions that are harmful in some sense. The case may be, for instance, an anorectic who states that she would rather die than gain weight, and therefore declines medical treatment that would otherwise have been good for her.

In such cases, it is important to consider whether the person is competent to make healthcare decisions. Here, being "competent" means that one can understand information about one's own situation, reason about the implications of the available care options, and so on. Sometimes anorectics are competent in this sense when they express wishes that are harmful. Therefore, psychiatrists, philosophers, and others, have analyzed such healthcare decisions in terms of authenticity. The question they raise is whether the person expresses authentic desires, in the sense that they correspond the person's "real" or "genuine" wishes.

An answer to the question whether a patient expresses authentic desires requires a theory about what it means for something to be authentic. In the first article in this thesis, I collect various theories that are intended to distinguish between authenticity and inauthenticity in a taxonomy that organizes the theories after category and class. Against this systematization, I argue that no category or class of authenticity theories gives rise to easily observable consequences. In short, I show that it is difficult—in the title of the article I claim too strongly that it is "impossible"—to determine whether someone else's desires are authentic.

This is the problem that I treat in the second article, which is also the most theoretical article in this thesis. In it, I build from a theory that explains authenticity in terms of *affirmative self-reflection*. According to this theory, a person's desire is authentic if she would approve of having the desire upon informed and critical self-reflection. One example that illustrates the theory is a heroinist who has the immediate desire to shoot heroin and also the reflected desire to lead a long and healthy life. The immediate desire is inauthentic, according to the theory, as it is not supported by the reflected desires.

The theory is difficult to put into practice, not least because it is difficult to collect all the reasons why a person has a certain desire and then put the person in a state of mind where she critically reflects upon those reasons and observe whether she, in this state of mind, approves of having the desire. It is an impractical theory. However, in the third article, I show that it nonetheless may give rise to practically observable consequences.

In the article, I reverse the central tenet of the theory. I argue that it is justified to judge that a person's desire is inauthentic if there is sufficient reason to believe that the person would *disapprove* of having the desire upon informed and critical self-reflection. With this formulation of the theory, it appears to be more practical. It is possible to identify factors indicating that a person would disapprove of her own desires. If enough such indicators are identified with sufficient reliability it is justified to make the judgment that her desires are inauthentic.

I suggest two such indicators in the article. The first indicator concerns the origin of the desire and whether it is normal to the person. If the desire is due to a medical disorder, for instance, that indicates that she would disapprove of having the desire if she came to know why she had it and reflected critically upon those causes. The second indicator concerns the coherence of the desire. If it does not fit with the desire-holder's practical identity, i.e., how she normally thinks and functions, that is a reason to believe that she would disapprove of having it upon informed and critical self-reflection.

My argument is not formulated in terms of necessary and sufficient conditions; I do not claim that judgments of inauthenticity are justified "if

and only if x and y." Instead, the arguments in my thesis should be understood against what philosophers usually call *reflective equilibrium*, which among other things means that the practical application of my theory requires contextual interpretations and substantial moral reasoning.

Furthermore, it should be noted that a judgment of inauthenticity does not suffice to justify paternalist interventions. That is, even if we are certain that a person's desires are inauthentic, that is not enough for it to be justified to, for instance, subject the person to compulsory care against her will. Paternalism must be independently justified, not least concerning the degree of confidence in judgments of inauthenticity and the reasonable proportion of the intervention.

In the third article of the thesis, I incorporate the theory which is spelled out in the second article into an already existing theory of personal autonomy, namely Tom L. Beauchamp and James F. Childress's (*Principles of Biomedical Ethics* 2013). I suggest that a patient's autonomy, i.e., her self-governance, can be analyzed in terms of authenticity and propose how this should be done. To demonstrate my further developed theory of autonomy, I construct a hypothetical case of anorexia nervosa. I use the authenticity-based concept of autonomy to analyze the anorectic's healthcare decisions and conclude that the analysis is fruitful and reliable. Therefore, I hold that the aim of the thesis has been met, although it is of course the critical reader who should make that judgment.

In the fourth article in this thesis I collect nine cases of authenticity-related problems in medical contexts. I apply the theory spelled out in the third article to the cases where this is possible. This shows that there are different kinds of authenticity problems; some concern the authenticity of decision-making, others concern possibly inauthentic mental states, and so on. Therefore, I argue that there is likely not one universal solution to authenticity-related problems in medical contexts, but rather various particular solutions to particular problems. The discussion shows that there is more theoretical work to be done in this field.

Populärvetenskaplig sammanfattning på svenska

Syftet med den här avhandlingen är att göra teoretiska ideal om autenticitet användbara i praktiken. Med autenticitet menas här att vara »genuin», »sann mot sig själv» eller liknande. Avhandlingen innehåller en »kappa» och fyra artiklar som är relaterade till autenticitet i en medicinsk kontext.

Problemen jag skriver om gäller autenticitet i beslutsfattandet, eller mer specifikt autentiska begär. Med »begär» menar jag och andra filosofer i det här sammanhanget det mest grundläggande elementet i vanligt beslutsfattande, en basal viljeriktning som påverkar en människas beslutsprocesser. Ett paradigmexempel som ofta används i den fackfilosofiska litteraturen för att illustrera autenticitetsrelaterade problem är personer som lider av anorexia nervosa och som fattar vårdbeslut som är skadliga i någon bemärkelse. Det kan till exempel gälla en anorektiker som säger att hon hellre dör än går upp i vikt, och som därför avstår från vårdinsatser som annars vore bra för henne.

När sådana fall uppstår är det viktigt att undersöka huruvida personen är beslutskompetent. Med »beslutskompetens» menas att man kan förstå information om ens egen situation, att man kan resonera kring innebörden av olika vårdalternativ, och så vidare. Ibland är anorektiker beslutskompetenta när de ger uttryck för viljor som är skadliga. Därför har psykiater, filosofer och andra analyserat sådana vårdbeslut i termer av autenticitet. Frågan de ställer sig är om personen ger uttryck för autentiska begär, i bemärkelsen att de stämmer överens med personens »riktiga» eller »genuina» önskemål.

För att besvara frågan om en patient ger uttryck för autentiska begär behövs en teori om vad det innebär för ett beslut att vara autentiskt. I avhandlingens första artikel samlar jag olika teorier som är avsedda att skilja mellan autenticitet och inautenticitet i en taxonomi som organiserar teorierna efter typ och klass. Mot denna systematisering argumenterar jag för att ingen typ eller klass av autenticitetsteorier ger upphov till enkelt observerbara konsekvenser. Kort sagt visar jag att det är svårt – i titeln hävdar jag för starkt att det är »omöjligt» – att avgöra huruvida någon annans begär är autentiska.

Det är detta problem jag tar mig an i den andra artikeln, som också är avhandlingens mest teoretiska. Jag utgår i den från en teori som förklarar autenticitet i termer av *självreflektivt bifall*. Enligt teorin är en människas begär autentiskt om hon skulle ge sitt stöd till begäret vid informerad och kritisk självreflektion. Ett exempel som illustrerar teorin är en heroinist som har det omedelbara begäret att skjuta heroin och samtidigt ett reflekterat begär efter ett långt och hälsosamt liv. Det omedelbara begäret är inautentiskt, enligt teorin, eftersom det inte har stöd av de reflekterade begären.

Teorin är svår att omsätta i praktiken, inte minst eftersom det är svårt att samla alla skäl till varför en person har ett visst begär och sedan försätta personen i ett sinnestillstånd så att hon reflekterar kritiskt över dessa skäl och sedan observera huruvida hon i detta tillstånd ger sitt stöd till begäret eller inte. Det är en opraktisk teori. Men i den tredje artikeln visar jag hur den trots allt kan ge upphov till praktiskt observerbara konsekvenser.

Det första jag gör i artikeln är att vända på teorin. Jag argumenterar för att det är rättfärdigat att bedöma att en persons begär är inautentiskt om det finns tillräckliga skäl att tro att personen skulle misstycka till att ha begäret vid en informerad och kritisk självreflektion. Med denna formulering av teorin tycks den bli mer praktisk – det är nämligen möjligt att identifiera sakliga faktorer som indikerar att någon skulle misstycka till sina egna begär. Om vi med tillräckligt hög tillförlitlighet identifierar tillräckligt starka sådana indikatorer är vi berättigade att fälla omdömet att personens begär är inautentiska.

Jag föreslår två sådana indikatorer i artikeln. Den första indikatorn är om begäret har uppstått till följd av orsaker som är onormala för personen, som till exempel en sjukdom. Den andra indikatorn är om begäret inte stämmer överens med personens praktiska identitet, alltså hur hon vanligtvis tänker och fungerar. Tillsammans ger de två indikatorerna skäl att tro att en persons begär är inautentiskt.

Min argumentation är inte uttryckt i termer av nödvändiga och tillräckliga villkor, jag hävdar inte att inautenticitetsomdömen är rättfärdigade »om och endast om x och y». I stället ska argumentationen förstås mot bakgrund av vad filosofer brukar kalla för *reflektivt ekvilibrium*,

vilket bland annat innebär att den praktiska applikationen av min teori kräver kontextuella tolkningar och substantiellt moraliskt resonerande. Det är i sammanhanget relevant att ett omdöme om inautenticitet inte är tillräckligt för att rättfärdiga paternalistiska interventioner. Det vill säga, även om vi är säkra på att en persons begär är inautentiska så räcker inte detta för att vi ska ha rätt att till exempel tvångsvårda personen mot hennes vilja. För att paternalism ska vara rättfärdigat krävs ytterligare stöd, inte minst vad gäller graden av säkerhet i inautenticitetsomdömet och den rimliga proportionaliteten i interventionen.

I avhandlingens tredje artikel inkorporerar jag teorin som stavas ut i den andra artikeln i en redan befintlig teori om personlig autonomi, nämligen den som Tom L. Beauchamp och James F. Childress har formulerat (*Principles of Biomedical Ethics* 2013). Jag föreslår att en patients autonomi, alltså hennes självbestämmande, kan analyseras i termer av autenticitet och föreslår hur detta ska genomföras. För att demonstrera mitt utvecklade autonomibegrepp konstruerar jag ett hypotetiskt fall av anorexia nervosa. Jag använder det autenticitetsbaserade autonomibegreppet för att analysera den hypotetiska anorektikerns vårdbeslut och drar slutsatsen att analysen är tillförlitlig.

I avhandlingens fjärde artikel samlar jag nio fall med autenticitetsrelaterade problem i medicinska sammanhang. Jag applicerar teorin som stavas ut i den tredje artikeln på de fall då detta är möjligt. Jag argumenterar för att det finns olika typer av autenticietsrelaterade problem: Vissa gäller autenticitet i beslutsfattandet, andra gäller möjligt inautentiska sinnestillstånd, och så vidare. Därför argumenterar jag också för att det troligtvis inte finns en enda universallösning på autenticitetsrelaterade problem i medicinska sammanhang, utan att det krävs flera olika lösningar för olika typer av problem. Diskussionen visar att det finns mer teoretiskt forskningsarbete att utföra i den här kontexten.

Delar av den svenska sammanfattningen är tidigare publicerad som Ahlin Marceta, J. (2018, 5 juli). Nytt forskningspapper om vad som rättfärdigar inautenticitetsomdömen [Blogginlägg]. Hämtad (2018-09-11) från https://jahlinmarceta.com/2018/07/05/nytt-forskningspapper-om-vad-som-rattfardigar-inautenticitetsomdomen/.

Theses in Philosophy from KTH Royal Institute of Technology

Below are listed all theses in philosophy that have been presented and defended at KTH Royal Institute of Technology since the inception of its Division of Philosophy in 2000. A licentiate thesis may be defended half-way between a master's degree and a doctor's degree. The content of a licentiate thesis could therefore be partially or wholly incorporated into a subsequent doctoral thesis by the same author. Some theses are freely available through the KTH Publication Database DiVA. The series has ISSN 1650-8831.

1. Martin Peterson, *Transformative Decision Rules and Axiomatic Arguments for the Principle of Maximizing Expected Utility*, Licentiate thesis, 2001.
2. Per Sandin, *The Precautionary Principle: From Theory to Practice*, Licentiate thesis, 2002.
3. Martin Peterson, *Transformative Decision Rules: Foundation and Applications*, Doctoral thesis, 2003.
4. Anders J. Persson, *Ethical Problems in Work and Working Environment Contexts*, Licentiate thesis, 2004.
5. Per Sandin, *Better Safe than Sorry: Applying Philosophical Methods to the Debate on Risk and the Precautionary Principle*, Doctoral thesis, 2004.
6. Barbro Björkman [now: Barbro Fröding], *Ethical Aspects of Owning Human Biological Material*, Licentiate thesis, 2005.
7. Eva Hedfors, *The Reading of Ludwik Fleck: Sources and Context*, Licentiate thesis, 2005.
8. Rikard Levin, *Uncertainty in Risk Assessment—Contents and Modes of Communication*, Licentiate thesis, 2005.
9. Elin Palm, *Ethical Aspects of Workplace Surveillance*, Licentiate thesis, 2005.

10. Jessica Nihlén Fahlquist, *Moral Responsibility in Traffic Safety and Public Health*, Licentiate thesis, 2005.
11. Karin Edvardsson, *How to Set Rational Environmental Goals: Theory and Applications*, Licentiate thesis, 2006.
12. Niklas Möller, *Safety and Decision-Making*, Licentiate thesis, 2006.
13. Per Wikman Svahn, *Ethical Aspects of Radiation Protection*, Licentiate thesis, 2006.
14. Hélène Hermansson, *Ethical Aspects of Risk Management*, Licentiate thesis, 2006.
15. Madeleine Hayenhjelm, *Trust, Risk and Vulnerability*, Licentiate thesis, 2006.
16. Holger Rosencrantz, *Goal-setting and Goal-achieving in Transport Policy*, Licentiate thesis, 2006.
17. Kalle Grill, *Anti-paternalism*, Licentiate thesis, 2006.
18. Jonas Clausen Mork, *Is it Safe? Safety Factor Reasoning in Policy Making under Uncertainty*, Licentiate thesis, 2006.
19. Anders J. Persson, *Workplace Ethics: Some Practical and Foundational Problems*, Doctoral thesis, 2006.
20. Eva Hedfors, *Reading Fleck: Questions on Philosophy and Science*, Doctoral thesis, 2006.
21. Nicolas Espinoza, *Incomparable Risks, Values and Preferences*, Licentiate thesis, 2006.
22. Mikael Dubois, *Prevention and Social Insurance—Conceptual and Ethical Aspects*, Licentiate thesis, 2007.
23. Birgitte Wandall, *Influences on Toxicological Risk Assessments*, Licentiate thesis, 2007.
24. Madeleine Hayenhjelm, *Trusting and Taking Risks: A Philosophical Inquiry*, Doctoral thesis, 2007.
25. Hélène Hermansson, *Rights at Risk—Ethical Issues in Risk Management*, Doctoral thesis, 2007.
26. Elin Palm, *The Ethics of Workplace Surveillance*, Doctoral thesis, 2008.
27. Jessica Nihlén Fahlquist, *Moral Responsibility and the Ethics of Traffic Safety*, Doctoral Thesis, 2008.

28. Barbro Björkman [now: Barbro Fröding], *Virtue Ethics, Bioethics, and the Ownership of Biological Material*, Doctoral thesis, 2008.
29. Karin Edvardsson Björnberg, *Rational Goal-Setting in Environmental Policy: Foundations and Applications*, Doctoral thesis, 2008.
30. Holger Rosencrantz, *Goal-Setting and the Logic of Transport Policy Decisions*, Doctoral thesis, 2009.
31. Kalle Grill, *Anti-paternalism and Public Health Policy*, Doctoral thesis, 2009.
32. Marion Godman, *Philosophical and Empirical Investigations in Nanoethics*, Licentiate thesis, 2009.
33. Niklas Möller, *Thick Concepts in Practice: Normative Aspects of Risk and Safety*, Doctoral thesis, 2009.
34. Johan E. Gustafsson, *Essays on Value, Preference, and Freedom*, Licentiate thesis, 2009.
35. Lars Lindblom, *The Employment Contract between Ethics and Economics*, Doctoral thesis, 2009.
36. Eduardo Fermé, *On the Logic of Theory Change: Extending the AGM Model*, Doctoral thesis, 2011.
37. Linda Johansson, *Robots and Moral Agency*, Licentiate thesis, 2011.
38. Johan E. Gustafsson, *Preference and Choice*, Doctoral thesis, 2011.
39. Per Norström, *Technology Education and Non-scientific Technological Knowledge*, Licentiate thesis, 2011.
40. Jonas Clausen Mork, *Dealing with Uncertainty*, Doctoral thesis, 2012.
41. Per Wikman-Svahn, *Ethical Aspects of Radiation Risk Management*, Doctoral thesis, 2012.
42. Karim Jebari, *Crucial Considerations: Essays on the Ethics of Emerging Technologies*, Licentiate thesis, 2012.
43. Linda Johansson, *Autonomous Systems in Society and War: Philosophical Inquiries*, Doctoral thesis, 2013.
44. Dan Munter, *Ethics at Work: Two Essays on the Firm's Moral Responsibilities towards Its Employees*, Licentiate thesis, 2013.

45. Sara Belfrage, *In the Name of Research: Essays on the Ethical Treatment of Human Research Subjects*, Doctoral thesis, 2014.
46. Patrik Baard, *Sustainable Goals: Feasible Paths to Desirable Long-Term Futures*, Licentiate thesis, 2014.
47. William Bülow, *Ethics of Imprisonment: Essays in Criminal Justice Ethics*, Licentiate thesis, 2014.
48. Per Norström, *Technological Knowledge and Technology Education*, Doctoral thesis, 2014.
49. Anna Stenkvist, *Pictures, Mathematics and Reality: Essays on Geometrical Representation and Mathematics Education*, Licentiate thesis, 2014.
50. Karim Jebari, *Human Enhancement and Technological Uncertainty: Essays on the Promise and Peril of Emerging Technology*, Doctoral thesis, 2014.
51. Mikael Dubois, *The Justification and Legitimacy of the Active Welfare State: Some Philosophical Aspects*, Doctoral thesis, 2015.
52. Payam Moula, *Ethical Aspects of Crop Biotechnology in Agriculture*, Licentiate thesis, 2015.
53. Jesper Jerkert, *Philosophical Issues in Medical Intervention Research*, Licentiate thesis, 2015.
54. Alexander Mebius, *Philosophical Controversies in the Evaluation of Medical Treatments: With a Focus on the Evidential Roles of Randomization and Mechanisms in Evidence-Based Medicine*, Doctoral thesis, 2015.
55. Patrik Baard, *Cautiously Utopian Goals: Philosophical Analyses of Climate Change Objectives and Sustainability Targets*, Doctoral thesis, 2016.
56. William Bülow, *Unfit to Live among Others: Essays on the Ethics of Imprisonment*, Doctoral thesis, 2017.
57. Björn Lundgren, *Semantic Information and Information Security: Definitional Issues*, Licentiate thesis, 2016.
58. Robert Erdeniz, *Military Operations Planning and Methodology: Thoughts on Military Problem-Solving*, Licentiate thesis, 2017.

59. Jesper Ahlin, *Personal Autonomy and Informed Consent: Conceptual and Normative Analyses*, Licentiate thesis, 2017.
60. Li Zhang, *On Non-Prioritized Multiple Belief Revision*, Doctoral thesis, 2018.
61. Björn Lundgren, *Information, Security, Privacy, and Anonymity: Definitional and Conceptual Issues*, Doctoral thesis, 2018.
62. Jesper Ahlin Marceta, *Authenticity in Bioethics: Bridging the Gap between Theory and Practice*, Doctoral thesis, 2019.