Personal Autonomy and Informed Consent: Conceptual and Normative Analyses

Jesper Ahlin

Licentiate thesis in philosophy
KTH Royal Institute of Technology
Stockholm, 2017

Ahlin, J.
Personal Autonomy and Informed Consent: Conceptual and Normative Analyses

Licentiate thesis KTH Royal Institute of Technology

ISBN 978-91-7729-495-5 ISSN 1650-8831

Typeset by the author, with help from Jesper Jerkert, using LaTeX and BibTeX. Printed by Universitetsservice US-AB, Drottning Kristinas väg 53B, 114 28 Stockholm.

I know what I know

Acknowledgments

I am grateful to Barbro Fröding, Sven Ove Hansson, and Niklas Juth for their supervision of this thesis, and to Gert Helgesson for his many useful comments on an earlier draft. Also, I thank William Bülow, Jesper Jerkert, Björn Lundgren, Payam Moula, and Maria Nordström, among others, for support and stimulating discussions. My thanks are extended to the Higher seminar at the Division of Philosophy, KTH Royal Institute of Technology, the Higher seminar at the Centre for Healthcare Ethics, Karolinska Institutet, and the VR/FORTE research program for providing opportunities to discuss and develop these ideas. All errors and mistakes are of course my own.

This research was supported by the Swedish Research Council (VR) 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.

Stockholm, August 2017 *Jesper Ahlin*

Abstract

This licentiate thesis is comprised of a "kappa" and two articles. The kappa includes an account of personal autonomy and informed consent, an explanation of how the concepts and articles relate to each other, and a summary in Swedish.

Article 1 treats one problem with the argument that a patient's consent to treatment is valid only if it is authentic, i.e., if it is "genuine," "truly her own," "not out of character," or similar. As interventions with a patient's life and liberties must be justified, the argument presupposes that the authenticity of desires can be reliably determined. If the status of a desire in terms of authenticity cannot be reliably determined, discarding the desire-holder's treatment decision on the basis that it is inauthentic is morally unjustified. In the article, I argue that no theory of authenticity that is present in the relevant literature can render reliably observable consequences. Therefore, the concept of authenticity, as it is understood in those theories, should not be part of informed consent practices.

Article 2 discusses the problem of what it is to consent or refuse voluntarily. In it, I argue that voluntariness should be more narrowly understood than what is common. My main point is that a conceptualization of voluntariness should be agent-centered, i.e., take into account the agent's view of her actions. Among other things, I argue that an action is non-voluntary only if the agent thinks of it as such when being coerced. This notion, which at first look may seem uncontroversial, entails the counterintuitive conclusion that an action can be voluntary although the agent has been manipulated or coerced into doing it. In defense of the notion, I argue that if the agent's point of view is not considered accordingly, describing her actions as non-voluntary can be alien to how she leads her life. There are other moral concepts available to describe what is wrong with manipulation and coercion, i.e., to make sense of the counterintuitive conclusion. Voluntariness should be reserved to fewer cases than what is commonly assumed.

Contents

Part I

Kappa

1. Introduction

Mrs. Parmelia Davis had suffered from epilepsy for many years. In 1905 her husband, Mr. Davis, arranged for Mrs. Davis to see Dr. Pratt, who was then considered an authority on epilepsy. Based on the prevailing medical view that in women there is a relationship between epilepsy and the uterus, Dr. Pratt told Mrs. Davis that he would have to operate on her womb. Mrs. Davis underwent a first surgery, in which Dr. Pratt repaired some lacerations of the uterine cervix. She was asked to return to the hospital one or two weeks later for further treatment of a minor irritation. Upon her second visit, Mrs. Davis was anesthetized while Dr. Pratt surgically removed her uterus and ovaries.

Later, in court, Dr. Pratt stated that he "did not deem [Mrs. Davis] worthy [of any explanation]," and that he had "calmly deceived the woman" (Katz 1984, p. 50). Through his attorney, Dr. Pratt advanced that by placing herself in his care for treatment without explicit limitations of his authority, Mrs. Davis had consented to "that he may perform such operation as in his best judgment is proper and essential to her welfare" (pp. 50–1). The court eventually ruled in favor of Mrs. Davis.[1]

Expressed with the conceptual apparatus of this thesis, Mrs. Davis did not give her informed consent to the medical intervention. As a point of departure, informed consent can be understood as shorthand for a patient's informed, voluntary, and competent consent to or refusal of treatment, which amounts to a valid authorization or refusal of a medical intervention (cf. Eyal 2012). This thesis includes two articles connected to informed consent.

The first article, titled "The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent," treats one problem with the argument that a patient's consent to treatment is valid only if it is authentic, i.e., if it is "genuine," "truly her own," "not out of character," or similar.[2] As interventions with a patient's life and liberties must be justified, the argument presupposes that the authenticity of desires can be reliably determined. If the status of a desire in terms of authenticity cannot be reliably determined, discarding the desire-holder's treatment decision on

[1] *Pratt v. Davis 1905* is discussed by, among others, Brennan (1991), Katz (1984), Millenson (2011), and Strong (2013). My account is a synthesis of theirs.

[2] Forthcoming in *Medicine, Health Care and Philosophy*: DOI: 10.1007/s11019-017-9783-0

the basis that it is inauthentic is morally unjustified. In the article, I argue that no theory of authenticity that is present in the relevant literature can render reliably observable consequences. Therefore, the concept of authenticity, as it is understood in those theories, should not be part of informed consent. In what follows, I call this article "the authenticity article," or variations thereof.

The second article, titled "Toward an Agent-Centered Theory of Voluntariness," discusses the problem of what it is to consent or refuse voluntarily. In it, I argue that voluntariness should be more narrowly understood than what is common. My main point is that a conceptualization of voluntariness should be agent-centered, i.e., take into account the agent's view of her actions. Among other things, I argue that an action is not non-voluntary unless the agent thinks of it as such at the time of acting. This notion, which at first look may seem uncontroversial, entails the counterintuitive conclusion that an action can be voluntary although the agent has been manipulated or coerced into doing it. In defense of the notion, I argue that if the agent's point of view is not considered accordingly, describing her actions as non-voluntary can be alien to how she leads her life. There are other moral concepts available to describe what is wrong with manipulation and coercion, i.e., to make sense of the counterintuitive conclusion. Voluntariness should be reserved to fewer cases than what is commonly assumed. I call this "the voluntariness article."

The above reveals a set of underlying moral problems. First, it is not obviously so that patients' desires should be included in a moral analysis of treatment decisions. For instance, it can be argued that someone else, such as the treating clinician, should make the decision that she finds most appropriate with regards to society's interests. Below, I present a brief account of the background of Western contemporary medical ethics, and thereby show that today's individualist paradigm is a historical novelty that should not be taken for granted. However, I will not discuss the moral value of individualism here.

Second, the proposition that discarding a patient's treatment decision is sometimes unjustified requires elaboration. For instance, it can be argued that overriding a patient's treatment decisions is justified if doing so leads to more well-being for her. This is commonly called *paternalism*; the practice of interfering with a person's self-governance (sometimes against her will) with reference to that it is, in some way, in that person's interest to be interfered with accordingly.

Many contemporary medical ethicists share a carefully restrictive attitude toward paternalism, and the main reason why is that personal autonomy is found valuable.

Personal autonomy is understood as being self-directed or self-governed; to live by one's own standards and be sufficiently independent from the influence of other persons or conditions. It is commonly viewed as a main moral guide in modern medical practice.

Third, autonomy is usually understood as the moral underpinning of informed consent. However, there is no real consensus regarding how personal autonomy should be understood in detail. Furthermore, the connection between personal autonomy and informed consent is neither immediate nor obvious, and it is not clear precisely how informed consent and its constitutive components should be formulated.

This thesis is intended to contribute to contemporary philosophy of personal autonomy and informed consent. The "kappa" is structured as follows. Section 2 presents some historical background to the contemporary philosophical conversation on informed consent. It is based on Anglo-Saxon medical history, centered around the concept of consent, and written from an internalist perspective, meaning that it does not take external factors such as cultural and political developments into consideration. Section 3 elaborates on personal autonomy, informed consent, their relationship, and how the articles included in this thesis are connected and contribute to contemporary theory. Section 4 concludes, and briefly discusses future research. Section 5 is a summary of the thesis in Swedish.

2. Some background notes

Pratt v. Davis 1905 helps to illustrate a shift of moral paradigm in health care. It is one of the first in a series of court rulings that together with cultural, political, and other developments during the 20th century, would lead up to the present standard requirement of informed consent in healthcare, aiming at protecting patients' autonomy.[3] 20th century health care is marked by a transformation in which patients have gained an unprecedented social power, and the development of institutional practices that enable their self-determination to a higher extent than ever before.

To a contemporary reader, it is common sense that consent between persons is morally important. The concept of consent is likely present in her notional world,

[3] The influence of The Nuremberg Code and the World Medical Association Declaration of Helsinki on this development cannot be overstated (see e.g. Lynøe 1999). However, court rulings have been more detailed in their ethical articulations.

and holds a privileged position in her arsenal of moral ideas. This has not always been the case. It is a recent development in medical ethics to take patients' consent into consideration, even more recent to assign to it a central role in behavioral policies in medicine. Historically, the main ethical concern in medicine has been the physicians' virtues, duties, and obligations (Jonsen 2000; Faden and Beauchamp 1986). That is, the ethics of treatment have gravitated around the treat-*er* and not the treat-*ee*. The 20th century moral individualism of medicine is a historical novelty.

The known history of Western medical ethics begins with Hellenic, Hellenistic, and Roman medicine. During this era, healing became detached from spiritual practices and was recognized as a domain of its own. It began to evolve into a rational enterprise based on empirical observation and logical reasoning. The Greek physician Hippocrates of Cos (460–370 BC) collected writings on the rational art of medicine. The Hippocratic Collection includes several treatises devoted to how physicians should behave. The earliest known statement about medical ethics is found in a passage in a book titled *Epidemics 1*, written by Hippocrates himself. The central dictum of the passage is *to do good or to do no harm*, later interpreted as *to help and not to harm*, or simply *do no harm* (Jonsen 2000, p. 2). Although first of its kind, and one of the most well-known moral dictums, it is not it but the Hippocratic Oath which has been most influential in the history of medical ethics. Those who swear the Oath commit to, among other things, keeping the sick "from harm and injustice," and to work "for the benefit of the sick" (*Hippocratic oath*). The commitments are unresponsive to the preferences of those who are treated.

Although the nature of medical ethics remained essentially unchanged for more than two thousand years, the content of medical practitioners' moral obligations shifted with social, cultural, and political conditions. One shift in moral content occurred in connection to the Black Death. When the plague hit Europe in the 14th century, many practitioners chose not to flee or hide but rather to try to treat those affected. Thus, the ethical debate advanced and began to incorporate questions of under what circumstances people who have medical skills have special obligations to serve their community; a topic which was later revived in the 1980's with the AIDS epidemic (Jonsen 2000, pp. 46–7).

In the beginning of the 19th century, Dr. Percival published an essay titled "Medical Ethics; or, a Code of Institutes and Precepts, Adapted to the Professional Conduct of Physicians and Surgeons." It is the first use of the term "medical ethics" in the literature of medical morality (ibid., p. 58). In the essay, Percival suggests that the

medical practitioner should have the "character of a gentleman," i.e., yet another moral maxim gravitating around the treat-*er* and not the treat-*ee* (ibid.).

Elsewhere, Percival took part in a debate that stretched over the entire 19th century; should physicians lie to their patients? The debate was due to an acknowledged dilemma between the physician's paternalist commitment to *do no harm*, or to *do good*, and the patient's right to be informed about her condition (so that she can make choices thereafter). It marks the first sign of moral individualism in medical ethics, since the debate invites the thought that it could be in the patient's interest to know the truth. However, it would take long before truth-telling became the norm in health care. A study of physicians' ethical views and attitudes conducted in the late 1960s found that "more than half the physicians (53%) thought that it was ethically appropriate for a physician not to tell a cancer patient that she had been enrolled in a double blind clinical trial of an experimental anticancer drug" (Faden and Beauchamp 1986, p. 89). Nonetheless, the 19th century debate about truth-telling reveals that ethical inquiry had begun to adhere to patients as individual subjects, possibly also with moral entitlements.

The debate, however, was not fully grounded in a genuine concern for patients' interests, but to a large part in treating physicians'. In the 19th century, legal institutions had emerged that could help patients hold physicians accountable for their conduct. Telling the truth to patients could sometimes minimize the physician's risk of later being accused of misconduct. This concern was not new to physicians. O'Shea accounts for a historical episode that forms one of the earliest reports of consent to treatment (2011, pp. 9–10): In late sixth century, the Byzantine emperor Justin II was gravely ill. He commanded his physicians to operate on him, but out of fear of being held responsible for his death they refused. Although the emperor promised they would not be punished, the physicians required that he personally handed them the scalpel as an explicit sign of his consent. Contrary to today's practice, the emperor's consent was intended to protect the consent-seeker and not the consent-giver. It was concerns of this kind that re-entered the ethical debate in the 19th century.

The mid-century physician Worthington Hooker belongs to the first champions of a genuine patient-oriented view in medical ethics. Hooker criticized the practice of deceiving patients, with the aim of protecting *them* and not the physicians (ibid., pp. 70–2). Thereby, an important step was taken toward today's autonomy-based consent, or the ideal practice of orienting health care around patients' self-determination. However, the most notable ethical developments thereafter were due to court rul-

ings rather than intellectual debate. During the 20th century, a series of legal cases, predominantly in North America and Great Britain, shaped the current doctrine of informed consent in medicine (Murray 1990). *Pratt v. Davis 1905* was one of the earliest such court rulings. When ruling in favor of Mrs. Davis, the court issued the following:

> [U]nder a free government at least, the free citizen's first and greatest right, which underlies all the others—the right to the inviolability of his person, in other words, his right to himself—is the subject of universal acquiescence, and this right necessarily forbids a physician or surgeon, however skillful or eminent … to violate without the permission the bodily integrity of his patient … and [to operate] on him without his consent or knowledge. (Katz 1984, p. 51)

The court's argument rests on *individual rights*. Together with other court rulings, it marks an important shift in the moral paradigm in medicine. Among the most noteworthy cases are *Schoendorff v. Society of New York Hospital 1914*, which made clear that the patient should be an active participant in the treatment decision process; *Bang v. Charles T. Miller Hospital 1955*, settling that patients are entitled to know inevitable risks or results of surgery; and *Reif v. Weinberger 1974*, establishing that any consent given under physical or mental duress is invalid (Murray 1990). These cases and others led the way toward today's concept of informed consent.

Thus, the individualism of today's medical ethics is a historical novelty. The development toward contemporary individualist theory is part of a larger societal context and is due to, among other things, interactions between religion, politics, law, moral philosophy, and science (cf. Siedentop 2014). Informed consent now governs behavioral policies in medicine in the Western world, but it should not be taken for granted.

3. Autonomy, informed consent, and the articles in this thesis

3.1. Personal autonomy and informed consent

Beauchamp and Childress's book *Principles of Biomedical Ethics* is arguably the most influential book in contemporary bioethics (1983; 2013). In it, autonomy is one

of four principles that in conjunction 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. Many bioethicists would argue that none of these principles takes precedence over the others. For instance, the case can be made that forced vaccination is morally acceptable if it is necessary to avoid a social catastrophe; the autonomy of those who are forced is then held as morally less significant than the harm avoided. However, the opposite position, i.e., that the principle of autonomy takes precedence over the others, can also be—and has been—defended:

> 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. (Gillon 2003, p. 310)

In any case, personal autonomy is central to contemporary medical ethics.

Informed consent aims at protecting patients' autonomy (albeit some have argued that it should have other aims, see e.g. O'Neill 2003). The articles in this thesis contribute to the theory of personal autonomy and informed consent in two essentially different ways, the discussion of which requires an elaboration of personal autonomy as a conceptual construct.

There is no consensus regarding how personal autonomy should be understood. Nonetheless, the "many faces of autonomy" may not be as numerable as some have suggested (cf. Taylor 2009, Ch. 2). It is generally held that autonomy, in the moral sense relevant to the present discussion, is a property that persons can enjoy to different degrees. Nonpersons, such as very small children, animals, and inanimate objects cannot be autonomous in this sense. Autonomy can be enjoyed to different degrees, meaning that it is not a binary concept; a person can be more or less autonomous, as well as non-autonomous and fully autonomous. And, 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 the 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, substantialists also take an interest in whether an agent's choices are self-supporting. To demonstrate, 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 choice. Substantialists, on the other hand, are concerned also with the fact that the victim chose not to leave her abusive partner. They may conclude that, as the victim's choice is self-injurious rather than self-supporting, it is non-autonomous. Proceduralists tend to accuse substantialists for unjustified paternalism, who tend to reply that proceduralists unwarrantedly ignore the social embeddedness of personhood. I will not engage in that debate here. In what continues, I will only treat issues in the procedural tradition, in line with the standard accounts in medical ethics.[4]

For the present purposes, there are three major ways in which personal autonomy is relevant to informed consent: 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 them. To illustrate, consider a patient who, due to a clinician's mistaken belief that she will hurt herself and others if left unconstrained, is strapped to her hospital bed. 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 in an attempt to burst 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 distort her view of herself and her surroundings. She might, for instance, hold a non-autonomous wish to hurt herself,

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

be encouraged by a fellow drug induced patient to non-autonomously decide to do so, and autonomously act upon those wishes and decisions.

It is first at this point in the inquiry into autonomy theory that the articles included in this thesis can be further elaborated on. The authenticity article concerns the autonomy of agents' inner lives, in this sense, while the voluntariness article mainly concerns that of their outer lives. Therefore, the two articles require their own discussions of how they are connected to personal autonomy and informed consent theory. Accordingly, the authenticity article and its connection to contemporary theory is discussed in section 3.2, while that of the voluntariness article is discussed in section 3.3.

3.2. The authenticity article

In the authenticity article, I present the hypothetical case of Anna, "a young and promising professional ballet dancer." Anna is known to love dancing more than anything else. However, after having suffered a serious injury, she refuses to consent to a minor medical intervention that would allow her to continue dancing. Anna is informed about her condition, competent to choose, and refuses voluntarily. Her decision is surprising. Anna's doctor begins to ponder the question of whether Anna's refusal is "genuine," i.e., *authentic*. If it is not authentic, her doctor thinks, perhaps Anna's refusal should be discarded and her "true wishes" adhered to through a forced medical intervention. That is, the doctor considers whether authenticity should be a criterion of informed consent.

Not surprisingly, it is through the concept of authenticity that the article is connected to personal autonomy and informed consent theory. Different philosophers whose theories of autonomy are mutually incompatible nonetheless share the view that authenticity is crucial to the autonomy of desires and wishes. Before continuing the discussion of the authenticity article, I will first account for so-called split-level theories of autonomy, and then Christman's theory of autonomy. These accounts give a brief introduction to how authenticity is treated in autonomy theory.

Split-level theories of autonomy build on an idea of how humans function psychologically. They assert that there are (at least) two levels of human desires, ordered as "higher" and "lower" desires (cf. Frankfurt 1971 and Dworkin 1988).[5] Desires on

[5] Split-level theories are not always formulated in terms of "desires," but rather in terms of "preferences," "wishes," etc. However, I take desires to be the basic element in preference forming and, thus, the most

the lower level are instinctive, or reactive; in short, non-reflected. A person may, for instance, have an immediate desire to inject heroin and enjoy its intoxicating ecstasy. Desires on the higher level are the opposite, i.e., reflected. The same person may have a reflected desire to live a long, healthy, and sober life. In split-level theories, a choice or an action that originates in desires on the lower level are non-autonomous if they are not in harmony with desires on the higher level. The person in the example acts non-autonomously if she injects heroin, as the action originates in desires that are inconsistent with her desire to live a long, healthy, and sober life. Thereby, in split-level theories, autonomy is a matter of (a sort of) inner coherence. This can be expressed as authenticity, so that authentic desires are such that would be endorsed by the desire-holder on a hypothetical higher level of reflection (Juth 2005, p. 145; Sjöstrand and Juth 2014, p. 120).

Christman argues that autonomy is a political value (2009, pp. 109ff). For political theory to be able to incorporate personal autonomy as a value, it requires of the theorist that she models "the self," or "the person," that embodies autonomy. In Christman's theory, persons are culturally and intertemporally situated beings; they are autonomous relative to their social sphere, and their history, present, and future. The theory takes into account the socially constructed nature of persons, and their complex cultural and social identities; their socio-historical selves. Two conditions are specified as critical to autonomy, namely competence and authenticity:

> Relative to some characteristic C, where C refers to basic organizing values and commitments, autonomy obtains if:
>
> (Basic Requirements – Competence):
>
> 1. The person is competent to effectively form intentions to act on the basis of C. That is, she enjoys the array of competences that are required for her to negotiate socially, bodily, affectively, and cognitively in ways necessary to form effective intentions on the basis of C;
> 2. The person has the general capacity to critically reflect on C and other basic motivating elements of her psychic and bodily make-up; and
>
> (Hypothetical Reflection – Authenticity):

basic way of articulating the concept of split-level autonomy theory.

> 3. Were the person to engage in sustained critical reflection on C over a variety of conditions in light of the historical processes (adequately described) that gave rise to C; and
> 4. She would not be alienated from C in the sense of feeling and judging that C cannot be sustained as part of an acceptable autobiographical narrative organized by her diachronic practical identity; and
> 5. The reflection being imagined is not constrained by reflection-distorting factors.
>
> (Christman 2009, p. 155)

Thereby, in Christman's theory, autonomous choices and actions are such that fit into, and further, a person's self-sustained autobiographical narrative.[6]

Thus, the doctor who considers whether authenticity should be a criterion of informed consent does so with the intention to foster Anna's autonomy. However, the question presupposes that the authenticity of decisions can be reliably determined; if it cannot, authenticity should not be part of informed consent, as interventions with patients' lives and liberties must be justified. Therefore, in the authenticity article, I investigate the problem of observing authenticity. I conclude that it is impossible to reliably determine the status of desires in terms of authenticity. This has been concluded before, though only of *some* theories of authenticity in *some* contexts (Sjöstrand and Juth 2014). However, I employ a methodology that allows me to draw the stronger conclusion that *no* theory of authenticity that is present in contemporary literature on personal autonomy (in the procedural tradition) can produce reliably observable consequences, and that the authenticity of desires therefore cannot be part of informed consent in *any* contexts.

To investigate the prospect of different theories of authenticity to produce reliably observable consequences I have developed a taxonomy of characteristics displayed by such theories. The taxonomy contributes to personal autonomy and informed consent theory by providing a conceptual apparatus that enables theorists to treat authenticity more systematically than before. In the taxonomy, which takes a three-by-two shape, theories of authenticity can display *sanctionist*, *originist*, and *coherentist*

[6] Christman summarizes his view in prose: "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).

characteristics. These are the categories of the taxonomy. In sanctionist theories, i.e., theories built on characteristics typical of sanctionist ideals, authenticity concerns the desire-holder's attitude toward 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. Furthermore, the characteristics can be either *cognitivist* or *non-cognitivist*. In cognitivist theories, authenticity is a matter of rational deliberation; non-cognitivist theories do not commit to that. These are the classes of the taxonomy. A theory can display characteristics from different categories (to different degrees; see footnote 6 in the article), but the classes are mutually exclusive so that a theory is either one or the other.

No category, class, or combination thereof, enables an observer with reasonable means to reliably determine the status of a desire in terms of authenticity. Therefore, it cannot be determined whether intervening with a patient's decision-making with reference to the authenticity of her desires furthers or restrains the patient's autonomy. In conclusion, authenticity should not be part of informed consent.

3.3. The voluntariness article

Suppose that a nurse serves a patient a cup of coffee, and that the patient proceeds to consume it. Now, suppose also that prior to this, the patient had expressed a wish not to be medicated, asked the nurse if there was medicine in the coffee, that the nurse assured her that there was not, though in fact there was. Was the patient medicated voluntarily?

I introduce this question in the voluntariness article as a point of departure to an analysis of contemporary accounts of voluntariness in personal autonomy and informed consent theory. In contrast to the problem of authenticity, voluntariness is mainly connected to the outer life of persons. Before discussing the voluntariness article, I will account for Beauchamp and Childress's theory of autonomy, and how informed consent as a more detailed construct connects to personal autonomy.

Beauchamp and Childress assume "that everyday choices of generally competent persons are autonomous," and construct a theory thereafter (2013, p. 104). In short, they argue that 1) intentional actions taken by 2) people with adequate understanding, who 3) are free from control exerted by external sources or internal states are autonomous (2013, pp. 104–5). I will elaborate briefly on these criteria.

The criterion of intentional action is binary, as an action either is or is not in-

tentionally executed. Intentional actions differ from accidental, as they correspond to the agent's conception of the act in question (although acts, due to unforeseen circumstances, may not always have the planned outcomes).

The criterion of understanding is one of degree. An agent must not have complete understanding of a situation to be able to act autonomously; adequate understanding suffices. Conditions that may limit adequate understanding include illness, irrationality, and immaturity.

The criterion of non-control is also one of degree. A person can be subject to the influence of others, for instance through coercion and manipulation, to an extent that robs the person of self-directedness. Such external control lessens the degree of autonomy with which the person acts. Internally controlling influences include debilitating disease, psychiatric disorders, and drug addiction (Beauchamp and Childress 2013, p. 138).

Lastly, as only competent persons can be autonomous, Beauchamp and Childress devise standards of incompetence that explain what makes a person incapable of autonomous action. These standards include the inability to express or communicate a choice, to understand one's situation, to give risk/benefit-related reasons, and an inability to reach a reasonable decision (ibid., pp. 118).

The above is an adequate account of how autonomy of actions is generally understood. In what follows, I give a more detailed account of informed consent, as it is closely connected to autonomy of actions.

Informed consent is a conceptual construction aiming at protecting patients' autonomy. By giving her informed consent to a medical intervention, a patient gives her valid authorization thereof. Beauchamp and Childress propose a detailed construct of informed consent:

I. Threshold elements (preconditions)

1. Competence (to understand and decide)

2. Voluntariness (in deciding)

II. Information elements

3. Disclosure (of material information)

4. Recommendation (of a plan)

5. Understanding (of 3 and 4)

III. Consent elements

6. Decision (in favor of a plan)
7. Authorization (of the chosen plan)

(Beauchamp and Childress 2013, p. 124; original emphasis)

Here, I will elaborate on the elements of competence, information, and voluntariness, in that order. In the present context, competence is an element that refers to a patient's capacity to make health care decisions. A patient is competent, or decisionally-capacitated, 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).

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). Thus, a patient can be competent to consent to treatment in the morning, but incompetent to consent to the same treatment in the evening. And, she can be competent to make one treatment decision, while at the same time being incompetent to make another. Likewise, a person can be competent to make health care decisions, but incompetent to make decisions regarding her personal finances. Therefore, it must be determined in each case against objective standards of competence whether the patient is competent to make the particular decision in question.

Some health care decisions are complex while others are simple. An assessment of whether to undergo a long and painful treatment or change one's way of life radically, is different from an assessment of whether to remove a cast or live with one's arm fixed in plaster (until, perhaps, the cast falls off by itself). The former decision is complex, the latter is not. Therefore, the level of evidence for determining competence varies (Beauchamp and Childress 2013, p. 120). Complex health care decisions require a higher degree of confidence in the patient's decision-making competence than simple decisions.

Furthermore, for a patient to give her valid authorization of a medical intervention she must be informed about her present condition, its causes, and an estimation of how it would develop if left untreated. She must be informed about what treatment options there are, their respective merits and flaws, associated risks and uncertainties, and an estimation of their prognoses of success. And, the patient must be informed

about the purpose of seeking her consent, and "the nature and limits of consent as an act of authorization" (ibid., p. 125).[7]

The information provided to the patient must be comprehensible. Medical treatment is often complicated. A patient must not be able to understand precisely how a certain medicine functions, and how the body responds to it, as such understanding sometimes presupposes vast knowledge of biology, chemistry, and human anatomy. Instead, the patient must be able to understand how a treatment decision is likely to affect her way of life. For instance, some treatments may necessitate regular clinical checkups that make it impossible for the patient to live abroad (where e.g. health care is inaccessible or too expensive) for longer periods at the time. It is crucial to informed consent that information is disclosed to the patient so that she is able to understand it. A patient cannot autonomously decide in favor of or against a medical intervention if the information on which she bases her decision is too complicated for her to understand.

Although this principle is phrased as if it was yet another requirement on the patient, it is in fact one on the information provider. That is, being able to understand information should not be understood as an element of decision-making capability, i.e., a property of the decision-maker, but as an element of communication, i.e., an obligation on the information provider, who is bound to deliver her message pedagogically. This may sometimes require of her that she uses communicative aids such as images, dolls, interactive computer projections, and so on.

Lastly, a patient's consent must be voluntary. I call the most influential theory of voluntariness the Voluntariness as Control (VaC) theory. It is supported by, among others, Beauchamp and Childress (ibid.), Nelson et al. (2011), and Wall (2001). According to the VaC theory, an action is voluntary if it is free from controlling influences. More elaborately, voluntariness is determined by two necessary and jointly sufficient conditions, namely the lack of manipulation and coercion. In the coffee example, the patient was manipulated into believing that there was no medicine in her coffee. She was therefore under someone else's control, and acted non-voluntarily. Likewise, had the nurse pointed a gun at the patient and said "drink your coffee or I will shoot you," the patient had been controlled by coercion and thus acted non-voluntarily.

In my view, the notion of non-control is flawed. The VaC theory (as well as other theories) makes too strong claims, so that voluntary actions are deemed non-

[7] If there is research involved, the patient must also be informed about, among other things, research agendas and procedures.

voluntary. I hold that an action is non-voluntary only if the agent is coerced and thinks of it as non-voluntary when being coerced. The patient in the coffee example took her medicine voluntarily (but unknowingly). The parenthesis is added to emphasize that there are other terms and concepts than voluntariness available to morally assess the practice of manipulating patients.

In the voluntariness article, I argue that a theory of voluntariness should take the agent's point of view into account. I ask the reader to suppose that an agent is faced with the same choice twice, that she does not think of them as non-voluntary in either case, and that she later finds out that in one case the factual basis of her choice was manipulated by some other agent (see "Manipulation and voluntariness" in the voluntariness article). A theory according to which the agent's choice was voluntary in one of the cases but not in the other neglects the agent's point of view, which I believe is a mistake; it does not orient the concept of voluntariness around the agent, i.e., is not agent-centered.[8]

However, the notion that actions are non-voluntary only if the agent thinks of it as such when being coerced entails the (presumably counterintuitive) consequence that actions can be coerced yet voluntary. Point a gun at a person and tell her to do as you say; if she does not think of her actions as non-voluntary at the time, they are not. I explain this consequence with that if a person is not engaged with a choice, i.e., if she does not *care*, coercion makes no difference to the voluntariness with which she acts. Furthermore, the concept of voluntariness is not essential for a moral assessment of the case. We could still explain why coercion is wrong by referring to individual rights, the wrongness of initiated aggression, and so on, and leave voluntariness aside until it is indispensable for moral analysis.

The conclusion that coercion only negates voluntariness if the agent cares about being coerced entails further questions. What is it to care accordingly? And, how does the notion relate to personal autonomy?

In the article, I introduce the pre-theoretical conception that thinking about an action as non-voluntary means experiencing an inner resistance to it; an intolerance, anti-attitude, or unwillingness. For the present purposes, this holds also as the meaning of what it is to care about being coerced, i.e., of being engaged with acts performed under coercion. An agent who does not care accordingly is indifferent. I find it worthwhile to consider the question of whether indifferent agents can act

[8] This explains the title of the article.

autonomously. First, coerced agents are not self-guided and do therefore not act autonomously, no matter if they care about or are indifferent toward the performed act. But, caring about and being indifferent toward acts in general seems to make a difference to autonomy as such. If an agent is indifferent toward an action, coerced or not, she is not exercising self-guidance.

4. Summary, conclusions, and future research

In-depth analyses of personal autonomy gains from a distinction between autonomous wishes and decisions, and autonomous acts. Informed consent, a conceptual construct aiming at protecting patients' autonomy, is constituted by elements targeting specific aspects of personal autonomy. Thereby, informed consent can be thought of as a practical operationalization of the complex moral value of being a self-guided person.

The two articles of this thesis are connected to personal autonomy and informed consent. Both are about how patients' autonomy can be strengthened, though in different ways. The authenticity article treats the inner life of agents, while the voluntariness article mainly treats their outer life. The authenticity article's conclusion is that although the authenticity of desires may be important to the autonomy with which a person makes health care decisions, it cannot be reliably determined whether a desire is authentic or not. Therefore, authenticity cannot be part of informed consent. The voluntariness article's conclusion is that voluntariness should be more narrowly understood than what is common. And, it suggests that it makes a difference to a person's autonomy whether she is engaged with her own actions.

In future research, I wish to further explore the thought that it matters to a person's autonomy whether she is engaged with her own actions. It shows resemblance to Immanuel Kant's notion that an autonomous person is one who purposefully subjects herself to the moral law, i.e., the categorical imperative (e.g. Kant 1996, pp. 81–6). In Kant's ethics, a person is autonomous only if she is actively engaged in how she leads her life (if she does her duty for duty's sake). Thus, contemporary bioethical discussions on personal autonomy and informed consent may have neglected a major theme in autonomy theory.

Furthermore, on several occasions I have found that my views necessitate a strong and coherent defense of moral individualism in bioethics. Some theorists

have proposed that the individualism of contemporary autonomy theory fails to incorporate important sociological and moral insights (see e.g. Fisher 2013, Osuji forthcoming, Ho 2008, Sutrop 2011, and Widdows 2011). I wish to explore their arguments in more detail; perhaps I have to change my views, or attempt to defend moral individualism in bioethics at greater length.

Also, in the authenticity article, I make the argument that interventions with a patient's life and liberties are unjustified if it cannot be reliably shown that her desires are inauthentic. This argument presupposes a strong view of justification, which can be contrasted with the weaker view that only *some* evidence is necessary for a paternalist intervention to be justified. Therefore, my conclusion in the article that the notion of authenticity should not be part of informed consent practices is not well-grounded. It requires a defense of an aspect of hard paternalism that I have yet to explore.

In addition to these normative and conceptual analyses, I have also taken an interest in the methods of applied ethics (see e.g. Archard 2016 and Wolff 2011). More specifically, I wish to study the methods of pluralist theories such as Beauchamp and Childress', which is usually understood as a "middle way" between "top-down" and "bottom-up" approaches to applied ethics (Beauchamp and Childress 1983; 2013). Pluralist normative theories such as theirs must allow for genuine conflicts between moral values, i.e., conflicts that cannot be solved *a priori*. Therefore, it seems to me that such theories should incorporate some idea of how to make moral judgments *a posteriori*. In future studies, I wish to explore whether the Aristotelian notion of *phronesis*, i.e., practical wisdom, could be fruitful as an approach in the methods of pluralist bioethics to solve moral dilemmas *a posteriori*.

However, although the above contributes (or would contribute) to personal autonomy and informed consent, the contribution is (or would be) largely theoretical. In the future, I wish to make more practical investigations of a normative character. I have two concrete projects in mind.

First, one of the most important distinctions in ethics is between facts and values. It has been shown that physicians who feel strongly for or against a treatment sometimes differ in their estimation of medical indication for this treatment, as compared to physicians who have no such sentiment (Björk et al. 2016). This may, by extension, affect the factual basis on which patients make treatment decisions and, thus, be of importance to their autonomy.

Second, in studies of immigrant patients, it has been shown that health care

consultations are more successful if there is a mutual linguistic and cultural understanding between patient and health care provider (Wiking 2012). I wish to study these potential problems and, if necessary, attempt to provide normative guidance with regards to them.

5. Swedish summary (sammanfattning på svenska)

Den här licentiatavhandlingen innehåller två artiklar om så kallat *informerat samtycke*. Informerat samtycke förstås ibland som informerat, frivilligt och kompetent samtycke, och är ett begrepp inom bioetiken som åsyftar en patients giltiga auktorisering av ett medicinskt ingrepp (eller en forskningsdeltagares auktorisering av vetenskapliga försök). Begreppets underliggande syfte är att försvara eller främja patientens *autonomi*, vilket i generella termer betyder »självstyre« eller »egenmakt«. I den här svenska sammanfattningen redogör jag kort för autonomibegreppet, informerat samtycke, avhandlingens två artiklar samt deras koppling till samtida teori.

Den första artikeln i den här avhandlingen rör autenticitet. Det kan argumenteras för att ett informerat samtycke är ogiltigt om det inte härrör från patientens »genuina« eller »egna« begär, det vill säga hennes *autentiska* begär. I artikeln »The Impossibility of Reliably Determining the Authenticity of Desires: Implications for Informed Consent« argumenterar jag för att oavsett om det stämmer eller inte så kan autenticitet i begären inte bli tillförlitligt observerat.[9] Autenticitet är en egenskap som, om den över huvud taget finns, inte kan mätas och utvärderas. Eftersom interpersonell moral kräver att medicinska interventioner rättfärdigas kan därför inte autenticitet ingå som kriterium i det informerade samtycket, även om vi på filosofisk väg skulle finna det önskvärt. Jag kallar den här artikeln för »autenticitetsartikeln«, eller variationer därav.

Den andra artikeln i den här avhandlingen handlar om frivillighet. De flesta teorier om frivillighet underlåter att ta agentens perspektiv i beaktande. Jag argumenterar i artikeln »Toward an Agent-Centered Theory of Voluntariness« för att en handling är ofrivillig endast om agenten tänker på den som sådan då hon blir tvingad att utföra den.[10] Det följer av detta att en handling kan vara frivilligt utförd även om agenten

[9] Publicerad online: Ahlin, J. (forthcoming) The impossibility of reliably determining the authenticity of desires: implications for informed consent. *Medicine, Health Care and Philosophy*. DOI: 10.1007/s11019-017-9783-0

[10] Skickad till tidskrift för referentgranskning.

utsätts för tvång, vilket enligt många är kontraintuitivt. Därför argumenterar jag för att det är viktigt för frivilligheten hos en handling att agenten är engagerad i den – om hon är oengagerad, eller med andra ord inte bryr sig om handlingen, så spelar det heller ingen roll för frivillighetens skull att hon blir tvingad till att utföra den. Det bör i sammanhanget även noteras att tvång kan fördömas moraliskt på andra sätt än med genom begreppet frivillighet. Vi kan till exempel säga att tvång utgör en rättighetskränkning, eller att det är odygdigt att nyttja tvång, med mera. Frivillighet bör förbli ett *smalt* eller *tunt* begrepp som endast är intressant att åberopa i undantagsfall, och inte per slentrian då det stämmer överens med våra intuitioner att göra så. Jag kallar den här artikeln för »frivillighetsartikeln«.

Ovanstående föranleder ett antal frågor av etisk karaktär. Viktigast för den här sammanfattningen är att båda artiklarna har som ändamål att stärka patienters autonomi, och att det inte är självklart hur informerat samtycke ska förstås. I vad som följer redogör jag i korthet för autonomibegreppet, det informerade samtycket, samt hur avhandlingens två artiklar är kopplade till samtida autonomi- och samtyckesteori.

Etymologiskt kan autonomibegreppet spåras till antiken. Idag – åtminstone i den nuvarande kontexten – menas med begreppet en persons självstyre eller egenmakt. Det finns både positiva och negativa aspekter av autonomi, i samma bemärkelse som positiv och negativ frihet. En autonom människa bär kapacitet att ägna sig åt kvalitativ självreflektion, hon kan utvärdera sina egna begär och värderingar och välja om, när och hur hon ska agera på dem. Detta är autonomi *positivt* uttryckt. En autonom människa lever inte heller under kontroll av andra agenter, influenser eller villkor. Detta är autonomi *negativt* uttryckt.

I de autonomiteorier som är relevanta i sammanhanget talas framför allt om autonoma önskningar, beslut och handlingar. Önskningar och beslut rör i första hand agenters inre liv medan handlingar rör deras yttre. Det inre livets autonomi har en koppling till informerat samtycke som åskådliggörs genom autenticitetsartikeln, medan det yttre livets autonomi har en koppling som visas genom frivillighetsartikeln.

En vanlig typ av autonomiteori som rör det inre livet kallas för *dualistiska* (»split-level«) teorier. Dualistiska teorier utgår från en viss idé om hur människan fungerar psykologiskt. Idén är att det finns (minst) två nivåer av mänskliga begär, ett »högre« och ett »lägre«. Begär på den lägre nivån är instinktiva, animaliska, eller reaktiva; i korthet oreflekterade. En person kan till exempel ha det omedelbara begäret att injicera heroin. Begär på den högre nivån är det motsatta, det vill säga reflekterade. Samma person kan ha det reflekterade begäret att leva ett långt, hälsosamt och nyktert

liv. I dualistiska teorier är ett val eller en handling som har sitt ursprung i de lägre begären icke-autonoma om de inte är i harmoni med begären på högre nivå. Personen i exemplet agerar icke-autonomt om hon injicerar heroin eftersom handlingen är i konflikt med hennes högre begär. Därmed kan sägas att dualistiska autonomiteorier handlar om inre, reflekterad koherens.

Det har argumenterats för att informerat samtycke borde innehålla ett autenticitetskriterium, vilket ytterligare skulle verka för att stärka patienters autonomi med avseende på det inre livet. I autenticitetsartikeln visar jag att det kanske är eftersträvansvärt i teorin, men att det inte kan genomföras i praktiken.

Enligt ett slags autonomiteori som rör det yttre livet är de flesta vardagliga handlingar som utförs av normalt kompetenta människor autonoma. Exempel på omständigheter som minskar graden av autonomi hos en handling är agentens oförmåga att förstå sin egen livssituation, eller handlingar som utförs på grund av manipulation eller tvång. Frivillighetsartikeln är avsedd att bidra till samtida teori med avseende på just det yttre livets autonomi. För att kunna placera detta bidrag i sitt rätta sammanhang ska jag först redogöra något mer utförligt för informerat samtycke som begrepp.

Informerat samtycke förstås ofta som en förkortning av en patients informerade, frivilliga och kompetenta samtycke. Med *informerat* samtycke menas att en patient är informerad om sitt medicinska tillstånd, dess orsaker och den förväntade utvecklingen om tillståndet förblir obehandlat. Vidare måste patienten vara informerad om vilka behandlingsalternativ som finns tillgängliga, deras respektive för- och nackdelar samt de risker och osäkerheter som är förknippade med de olika alternativen. Slutligen måste patienten vara informerad om samtyckets syfte, natur och gränser.

Ett problem vad gäller detta kriterium är att informationen måste vara förståelig. Detta kan till synes vara ett problem på patientsidan av samtycket, men är i själva verket ett problem på den behandlande sidan: det åligger agenten som informerar patienten att göra informationen förståelig. Med andra ord har problemen som just nämnts inte med patientens kompetens att göra, utan gäller kommunikation. Det kan ibland vara moraliskt ålagt den som behandlar patienten att använda pedagogiska hjälpmedel som bilder, dockor eller interaktiva dataprojektioner för att förmedla information till denne.

Vidare måste ett samtycke vara *kompetent* för att vara giltigt. Med kompetens menas här patientens kapacitet att fatta beslut om sin egen vård. En patient är kompetent om hon kan förstå information, uppskatta hur den berör henne och resonera

kring informationen med hänsyn till hennes egna värderingar och preferenser. Dessa kompetenser förutsätter även andra. Till exempel krävs det att patienten kan tänka kritiskt på sig själv som ett intertemporalt subjekt, vilket är en förmåga som ofta saknas barn (och andra).

Det bör noteras att kompetens är ett beslutsrelativt begrepp. En patient kan till exempel vara kompetent att fatta ett beslut på morgonen men på kvällen vara inkompetent att fatta samma beslut. Hon kan också vara kompetent att fatta ett visst beslut men vid samma tidpunkt vara inkompetent att fatta ett annat, liksom att hon kan vara kompetent att fatta beslut som rör hennes vård men inkompetent att fatta beslut som rör hennes privatekonomi.

Slutligen måste patientens samtycke vara *frivilligt*. Det finns flera teorier om hur frivillighet ska förstås, men den fortsatta diskussionen kräver några avgränsningar. För det första saknar vi på svenska en motsvarighet till engelskans »involuntary«. Därmed måste engelskans tre begrepp – »voluntary«, »non-voluntary« och »involuntary« – reduceras till två. Med »involuntary« menas vanligen sådana handlingar som en agent inte i någon egentlig mening *utför*, utan som snarare *händer henne*. Till exempel är blinkningar och nysningar inte frivilliga i någon moraliskt meningsfull bemärkelse, utan sorterar under engelskans »involuntary«. Svenskans begreppspar *frivillig* och *ofrivillig* handlar därmed om en agents utförda handlingar, i kontrast till sådana som på svenska kanske bäst beskrivs som *reflexer*.

För det andra gäller frivillighet här endast val och handlingar som utförs människor emellan. Begreppet kan annars också användas för att illustrera någonting metafysiskt, som i »fri vilja«, samt val och handlingar som utförs på grund av naturkrafter, som när en skeppare »ofrivilligt« kastar sin last överbord för att rädda sitt skepp från att kapsejsa.

Enligt de flesta teorier om frivillighet upphäver tvång frivillighet. Det vill säga, om en person tvingas att utföra en handling så är handlingen av den anledningen alltid ofrivillig. En teori som inte delar denna idé – och som jag, för tydlighets skull, inte stöder – är att handlingar är ofrivilliga om och endast om de utförs eftersom handlingsalternativen som agenten tror sig ha är oacceptabla enligt en objektiv standard om välbefinnande. Som undantag ges att en handling är frivillig om det enda acceptabla alternativet är ett sådant som agenten uppskattar så mycket att hon utför den av den anledningen. Det betyder att om *B* tvingar *A* att utföra handling φ, men *A* uppskattar φ så mycket att hon väljer att utföra φ av den anledningen, så utför *A* φ frivilligt. Med andra ord upphäver tvång enligt teorin inte alltid frivillighet.

Enligt den mest inflytelserika teorin om frivillighet handlar en agent frivilligt så länge hon vill utföra handlingen utan att vara utsatt för kontroll av någon annan agent eller (socialt) villkor. Med kontroll menas här framför allt manipulation och tvång, det vill säga att en agent kan vara kontrollerad antingen genom att vara manipulerad eller tvingad. Jag placerar denna teori centralt i frivillighetsartikeln.

Mitt argument i frivillighetsartikeln bygger på att en handling är ofrivillig endast om agenten tänker på den som sådan när hon blir tvingad att utföra den, vilket har som följd att handlingar kan vara manipulerade eller tvingade men ändå frivilliga. Det är avgörande om agenten är engagerad i den – om hon är oengagerad, eller med andra ord inte bryr sig om handlingen, så spelar det heller ingen roll för frivillighetens skull att hon blir tvingad till att utföra den.

Detta leder till tanken att det kanske är avgörande också för autonomin att en person är engagerad i sina egna beslut. Om hon är oengagerad, eller inte bryr sig, tycks hon ju inte utöva självstyre. Jag föreslår i avhandlingen att denna möjlighet bör utforskas vidare. I avhandlingen diskuterar jag också andra möjliga framtida studier. Jag förlitar mig till exempel i båda avhandlingens artiklar på en moralisk individualism, men ger inget utförligt försvar av individualismen. Likaså förlitar jag mig i autenticitetsartikeln på en strikt paternalism som jag lämnar oförsvarad.

Jag har även som ambition att i framtiden göra mer tillämpade utredningar, och har två sådana projekt i åtanke. Det första gäller den inom etiken välkända distinktionen mellan fakta och värderingar. Studier har visat att läkare som har en stark personlig preferens för en viss behandling tenderar att göra annorlunda riskvärderingar kring denna än vad läkare som saknar en sådan personlig preferens gör. Detta kan i förlängningen komma att påverka den faktabas som patienter fattar sina vårdbeslut mot, och därmed ha betydelse för patienters autonomi. Det andra projektet bygger på studier som visar att vårdkonsultationer är mer framgångsrika om vårdgivaren och vårdtagaren delar en ömsesidig lingvistisk och kulturell förståelse. Jag planerar att studera dessa fall närmre och, om det är nödvändigt, försöka att ge normativ vägledning i dem.

Part II

Included 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 protect patients with regards

[†] Published online: Ahlin, J. (forthcoming) The impossibility of reliably determining the authenticity of desires: implications for informed consent. *Medicine, Health Care and Philosophy*. DOI: 10.1007/s11019-017-9783-0

[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).

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 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

[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).

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. 2006a, 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 genuine," "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

[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.

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 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:

> 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… (Sjöstrand and Juth 2014, p. 121)

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.

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

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 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

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

[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.

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. 10–7; 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 be inclined to approve of having that desire if she came to know why she has it" (Juth 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" (Sjöstrand and Juth 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 (ibid., p. 20; 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]" (Tan et al. 2006b, p. 20).[7] Similarly, but more elaborately, Elster

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

argues when writing about the rationality of desires that 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; 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" (ibid., 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" (Meyers 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" (Meyers 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" (Meyers 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" (Christman 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:

> 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. (Miller 1981, p. 24)

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:

> 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. (Sjöstrand and Juth 2014, p. 121)

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 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.

[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.

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 over-

come, 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.

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.

Toward an Agent-Centered Theory of Voluntariness*

Abstract: The concept of voluntariness is central to informed consent and personal autonomy, yet it has been underexplored by bioethicists. There are various theories intended to explain voluntary choice and action. None is fully agent-centered, in the sense that the conceptualization of voluntariness takes into account the agent's views of her decisions and actions. An agent-centered theory of voluntariness would promote analytical precision, and foster autonomy in healthcare and research practices. According to the most influential bioethical theory of voluntariness, here called the Voluntariness as Control theory, an action is non-voluntary if the agent is controlled by external influences. The theory is critically discussed from an agent-centered perspective, and a new conceptualization of voluntariness is proposed.

Keywords: voluntariness, autonomy, informed consent, agent-centered, bioethics

Introduction

Voluntariness is a central concept in bioethics. It is most commonly associated with informed consent, i.e., an agent's valid authorization of a medical intervention or research participation. Consent to such interventions or practices is invalid if it is not voluntary. However, the concept of voluntariness has been underexplored by bioethicists. Multiple and mutually incompatible theoretical accounts are present in the literature on informed consent. Although one type of conceptualization has been more influential than others, no single tradition or school of thought overshadows the others completely. This article is intended to stimulate bioethicist debate on voluntariness; its conditions, and role.

No theory of voluntariness that is present in the contemporary bioethical literature is agent-centered, i.e., takes the agent's perspective into account. Here, it is argued

* Submitted for review.

that, to foster personal autonomy, a conceptualization of voluntariness should be aligned with how the agent actually leads her life and acknowledge whether she is engaged with her decisions and actions. It is proposed that an action should be understood as non-voluntary if and only if it is coerced, and if the agent thinks of the action as non-voluntary when being coerced to perform it.

The article is structured as follows. First, I introduce three hypothetical cases that demonstrate the ambiguities of voluntariness as a pre-theoretical construct. Then, I outline five theories of voluntariness that are present in contemporary bioethical literature. The most influential, here called the Voluntariness as Control theory, is the main target of the subsequent arguments. In the sections that follow, I use a set of examples to show how the Voluntariness as Control theory is flawed, and why it is important that a theory of voluntariness is agent-centered. Lastly, I defend my proposed conceptualization from some expected objections, and put the theory at work by applying it to hypothetical cases.

Contemporary theories of voluntariness

When voluntariness is relevant

Consider these three hypothetical cases in which the concept of voluntariness could be invoked. First, suppose that a nurse serves a patient a cup of coffee. The patient, who does not wish to take any medicine, asks whether there is anything but coffee in the cup. It is only coffee, the nurse assures the patient, knowing that she does not wish to take any medicine. This continues routinely for a month or two. The nurse serves the patient coffee, persistently maintaining that there is nothing but coffee in the cup, and the patient consumes it. One day, the patient learns that there has been medicine dissolved in every cup of coffee that the nurse has served her. Did the patient take the medicine voluntarily?

Now, suppose that a young pregnant woman tells her doctor that her parents, her partner, her partner's parents, and all her friends think that she should have an abortion. She quotes them saying: "You are too young. You should finish school, get a job, marry, and then perhaps think about having children." When her doctor asks the patient what *she* wants, she repeats in first person: "I am too young. I should finish school, get a job, marry, and then perhaps think about having children." It is

clear to the doctor that the patient does not follow her own wishes, but others'.[1] If she would go through the procedure, would it be voluntarily?

Lastly, suppose that a patient has lived with a severe depression for a long time. She also suffers from some physical disease, the cure of which requires that the patient undergoes treatment. The patient refuses, and expresses a wish to die. Does she refuse treatment voluntarily? If her doctors enrolled the patient in a study to learn about the process of dying from that disease, would the patient's consent to participate be valid?

In neither case is the answer obvious. They are—supposedly—examples of manipulation, social pressure, and the potential lack of decision-making capacity, which all require careful moral analysis. However, it is not unquestionably clear whether the concept of voluntariness should be invoked to express whatever it is that is morally problematic with the cases. Conceptual ambiguities render it vague how the concept of voluntariness should be applied in healthcare and research. In what follows, I will briefly account for some conceptualizations of voluntariness that have been proposed in bioethical literature. Some demarcations are necessary, however.

First, some theorists wish to distinguish between voluntary choice and voluntary action. However, in the bioethical literature on voluntariness, "choice" and "action" are more often than not used indiscriminately. And, I fail to see why a distinction would be necessary. In the present context, the concept of voluntariness—disregarding of how it is understood—applies to informed *consent*. Consent could also be conceptualized both as choice and as action, so that consenting to something is "to decide in favor of that something," or similar, or "to perform a speech act that marks an approval of that something," or similar. Voluntariness would apply alike in either way; the agent consents voluntarily or non-voluntarily, disregarding of whether consent is understood as choice or as action. Accordingly, "choosing" and "acting" voluntarily are here merely proxies for the overarching "consenting" voluntarily, and will be used interchangeably.

Furthermore, in the present context, voluntariness is an interpersonal concept (cf. Nozick 1974, p. 262). Since ancient times, voluntariness has been discussed in wider contexts. Aristotle, for instance, questioned whether a sailor who throws his cargo overboard to save his ship from capsizing in a storm does so voluntarily (Aristotle. 2014, 1110a5–10). Facts of nature that forces us to make choices like the

[1] Whether observers really can determine the authenticity of others' desires is a whole other issue (see Ahlin forthcoming).

sailor's can be analyzed in terms of voluntariness, although it is not relevant for the present interpersonal purposes.

Moreover, doing something *non-voluntarily* is different from doing something *involuntarily*. While it is yet to be established what the former means, doing something involuntarily means that the action is not causally linked to the agent's decisional mechanisms. In short, reflexes such as blinking and sneezing are involuntary actions. It could also be argued that they are not actions at all, but I will not explore that possibility further here. It suffices to say that in the following, involuntariness is not of interest, why I will only treat non-voluntariness (and voluntariness).

Finally, it has been argued that the individualist framework of personal autonomy and informed consent is inappropriate and should give way to a structuralist narrative that better captures agents' lived experiences by acknowledging patterns of "class, race, ethnicity, sex, and other demographic characteristics" (Fisher 2013, pp. 256–60).[2] Others have suggested that individualism "abstracts from an interconnected social reality" and "overlooks the fact the people often base the social relationships on relational care" (Osuji forthcoming). Such critique marks a distinction between individualism and non-individualism that, in turn, entails a complete separation between different schools of bioethical thought. Therefore, I note that a defense of individualism *per se* would be desirable, although that is beyond the scope of the present purposes. Non-individualist theories of voluntariness will not be taken into further consideration here.

Theories of voluntariness

Most accounts of voluntariness share the notion that coercion always renders actions non-voluntary. Hyman discusses (but dismisses) an alternative view: "… it is plausible that threats, however terrible, change the incentives associated with a course of action, just as offers and bribes do, so that they *influence* but do not *abolish* choice, and therefore do not negate voluntariness" (Hyman 2013, p. 691). Olsaretti has developed a theory of voluntariness that does not build on coercion. In it, a choice is non-voluntary if it is made because the alternatives that the chooser believes she faces are unacceptable according to an objective standard of well-being (1998, 2008; see also Colburn 2008). That is, if *A* is coerced to do φ, but still has the acceptable

[2] See also chapter 4 in Bowman et al. (2012) for an analysis of consent in medical settings resting on power structures and authority.

alternative of doing δ, then she does φ voluntarily. Furthermore, Olsaretti holds that choice is voluntary if the only acceptable alternative is one that the agent likes so much that she chooses it because of that. So, if *A* is coerced to do φ, but likes φ so much that she chooses to do it because of that, then *A* does φ voluntarily.[3]

According to Hyman, the concept of voluntariness must be understood in light of its function, which is to "inform the appraisal of individual conduct," in particular "the assessment of innocence and guilt" (2013, p. 685). Hyman suggests that voluntariness is formed by negation, and provides the rough definition that "a certain thing is done voluntarily if, and only if, it is *not* done out of ignorance or compulsion" (2013, p. 685). He tracks the historical roots of his theory back to the antics, noting that ignorance is identified as an exculpation already in the *Iliad* (2013, p. 686).

Nozick proposes that non-rights-violating actions are voluntarily performed. The theory builds on the so-called "self-ownership thesis," i.e., the normative proposition that every individual person has an exclusive right to her own body and mind that amounts to a property in her self (Nozick 1974, p. ix; see also Ch. 1–2). As full property rights apply in Nozick's political libertarianism, the self-ownership thesis entails that no person is morally entitled to intervene with anyone else's person or project. Intervening (without permission) is a rights violation. In Nozick's theory, all rights-violating actions, and only rights-violating actions, negate voluntariness.

I call the theory which has been most influential in bioethics the Voluntariness as Control (VaC) theory. It is supported by, among others, Beauchamp and Childress (2013) and Nelson et al. (2011). According to the VaC theory, an action is voluntary if it is free from controlling influences. In the coffee example, the patient was manipulated into believing that there was no medicine in her coffee. She was therefore under someone else's control, and acted non-voluntarily. Likewise, had the nurse pointed a gun at the patient and said "drink your coffee or I will shoot you," the patient had been controlled by coercion and thus acted non-voluntarily. As the VaC theory is the most influential in bioethics, it will be the main target for the arguments in the following sections. Therefore, it deserves some elaboration.

The VaC theory

In their prominent book *Principles of Biomedical Ethics*, Beauchamp and Childress adhere to the VaC theory. They outline their version of the theory by writing that "a

[3] See Wertheimer (2012) for a critical discussion of Olsaretti's theory.

person acts voluntarily if he or she wills the action without being under the control of another person or condition" (2013, pp. 137–40). Similarly, Appelbaum et al. argue that voluntariness depends on "the extent to which subjects are actually exposed to external, intentional, and illegitimate influences that causally impact their decisions" (2009, p. 37). 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.[4] Voluntary action, they argue, should be understood in terms of the two necessary and jointly sufficient conditions of intentional action and the absence of controlling influences. 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.[5]

Examples of controlling influences in the broad sense include offers of payment, threats, education, deceit, manipulative advertising, emotional appeals, and the like (ibid., 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" (ibid., 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.[6]

Furthermore, a person can be controlled through coercion. Coercion is 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

[4] Their account forms the basis of at least one empirical study of voluntariness; Miller et al. (2011).

[5] According to Nelson et al., the theory of voluntariness as degree of control was first introduced as a detailed construct 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).

[6] 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 (2011, p. 7).

another person" (Nelson et al. 2011, p. 7). True coercion by threat "requires that a credible and intended threat disrupts and reorders a person's self-directed course of action" (ibid., p. 8).[7]

According to the VaC theory, the patient in the coffee example took her medicine non-voluntarily, at least given that someone—presumably the nurse—intentionally arranged the whole situation. If there had been a mix-up in the hospital kitchen so that the patient was served coffee intended to be served to some other patient who, for some reason, wished to have his or her medicine dissolved in coffee, the patient would not have been intentionally controlled by someone else. VaC theorists would then not deem the patient's actions non-voluntary. However, the opposite conclusion could also be drawn, if a distinction is introduced between control *negatively* and control *positively*, in a sense corresponding to negative and positive liberty.

The above account of control as non-manipulation and non-coercion is an account of control negatively construed, that is, acting and choosing without being controlled by some other person or condition. Being in control positively is something else. Nelson et al. write that the condition of noncontrol "is not sufficient for voluntary action, because an agent *in control* is likewise essential," and analyze "this agency" as "*capacity* for intentional action" (2011, p. 10). Similarly, Wall writes that control requires that one is "able to follow through and perform in ways that facilitate the realization of what is sought," which appears to be the idea of being able to direct one's behavior in accordance with one's intentions (2001, pp. 130–1). This is also how Beauchamp indicates that he understands control (Beauchamp 2010, p. 69).

The patient who, due to a mix-up in the hospital kitchen, is mistakenly served coffee with medicine dissolved in it is not controlled by some other person or condition. Yet, she lacks control positively construed; she is not in control over her medication. Therefore, a VaC theorist could argue that she does not take the medicine voluntarily.

Control positively construed, however, shows too much resemblance to the concept of personal autonomy. It obscures the important distinction between doing something *voluntarily* and doing something *autonomously*.

Beauchamp and Childress describe autonomy "in terms of normal choosers who act (1) intentionally, (2) with understanding, and (3) without controlling influences that determine their action" (2013, pp. 102–4). Taylor writes that "a person is autonomous if he guides and directs his decisions and actions in the light of his own

[7] This conceptual framework was first introduced by Nozick (1969).

desires and values, free from the interference of others" (Taylor 2009, p. xiii). And, Christman understands it similarly:

> In general, autonomy is meant to manifest self-government, the ability of the person to guide her life from her own perspective rather than be manipulated by others or be forced into a particular path by surreptitious or irresistible forces. (Christman 2009, p. 134)

Control positively construed is almost or completely identical to personal autonomy, as the concept is understood both generally in the field and specifically by some of the most prominent autonomy theorists. It is unclear what VaC theorists think of this resemblance. However, the distinction between voluntariness and autonomy should be upheld. It is important not least because voluntariness is partly constitutive of autonomy; an act can be voluntary and non-autonomous but not non-voluntary and autonomous. Therefore, I will not take the possibility of control positively construed into further consideration in this article.

An agent-centered theory of voluntariness

Manipulation and voluntariness

The VaC theorists' claim that manipulation negates voluntariness is too strong. There might be a connection between the factual basis for a choice or an act and the voluntariness with which it is performed, but the VaC theory overstates that connection. The theory renders inaccurate act descriptions, and thereby fails to be aligned with how agents lead their lives. Consider the following hypothetical examples of manipulated agents whose choices it would be misguided to describe as non-voluntary.

Suppose that a patient correctly believes that she has treatment options *A*, *B*, and *C* to choose from. She is completely indifferent with regards to how preferable they are; it does not matter at all to the agent which treatment she receives. The hospital has financial interests in patients choosing *non-A*, and the patient's doctor has been instructed to falsely tell patients that *A* is unavailable. The doctor complies, and based on the doctor's lie the patient chooses *B*.

It is an overstatement to claim that the patient chose *B* non-voluntarily. The patient could live her whole life not knowing that she was once deprived of option *A*. In her mind, the choice stood between the equally good *B* and *C*. Describing her

treatment decision as non-voluntary is alien to how she leads her life. If the patient were to find out about the truth, she might begin to think of option *A* as somewhat better than *B*; an irrational but understandable psychological reaction, similar to what behavioral economists call "loss aversion" (cf. *Loss aversion*). Nonetheless, it would be inaccurate of her to suddenly begin to describe a treatment choice which she has always been satisfied with as non-voluntary.

Yet, VaC theorists would argue just that. The patient did not "will the action without being under the control of another person or condition," as Beauchamp and Childress might wish to put it (see above). Likewise, Nelson et al. could point out that the patient did not believe that *B* is the best choice merely due to "the merit of reasons proposed," but also due to the doctor's lie, and is therefore choosing under controlling influences, i.e., non-voluntarily.

Their act description is uninformative—perhaps even deceptive, as it triggers certain intuitions that should not be triggered here. Certainly, even more so as the bioethicist moral vocabulary includes concepts such as "deceit," "obligations," and "rights," which are more suitable than "voluntariness" to apply here, while being sufficiently accurate to explain the ethics of the case. Given that the patient was wronged in some way, the concept of voluntariness is not indispensable for a description of precisely how she was wronged. In this case, it suffices to say that the doctor lied to her, which is a breach of confidence and a restriction of the autonomy with which the patient made her treatment decision.

Consider now an example in which a patient is not robbed of an option. Instead, an option is added to her range of available choices, and the patient is lied to concerning the new option's features.

A patient at a medical university correctly believes that she has treatment options *A*, *B*, and *C* to choose from. She ranks the alternatives in that order. Now, her doctor, who is also a researcher, informs the patient of an experimental treatment alternative *A+*. The doctor suggests that *A+* is superior to the other options, but the truth is that there is no evidence supporting the doctor's claims. Based on the doctor's lie, the patient chooses option *A+* and suffers painful consequences from medical maltreatment.

Again, VaC theorists would argue that the patient stood under the doctor's control, and thereby chose non-voluntarily. However, at least while there are other relevant moral concepts available, it is an imprecise conviction.

Describing the patient's choice as non-voluntary is alien to how she leads her life.

If she were to find about the truth, she would likely rank *A+* far below *A*, *B*, and *C*. Nonetheless, it would be inaccurate of her to suddenly begin to describe a treatment that she was misled to value incorrectly as non-voluntary first when, and because, she learns its true value.

Furthermore, the patient made what she thought of at the time as a conscious and self-determining choice. She was presented with a list of options, ranked them per her own preferences, decided in favor of one of them, and had her wishes adhered to. Describing her choice as non-voluntary is to belittle her role as a rational decision-maker, as the choice of terms indicates passivity, i.e., non-agency. It is to degrade the patient completely from person to prop on the scene of events.

Lastly, holding that the patient chose *A+* non-voluntarily is inflationary. It deprives the words "non-voluntary" of an intensity that we should strive to retain to be able to vividly and accurately depict gun-to-the-head scenarios that are sometimes a reality.

It is out of place to say that the patients in these examples chose non-voluntarily. Doing so adds little of value to a description of their actions, particularly in relation to the precision that could be attained by applying the analytically superior concept of autonomy to the cases: The patients chose voluntarily but—for reasons unrelated to voluntariness—non-autonomously. A theory that renders the patients' actions non-voluntary is suboptimal with regards to fostering personal autonomy and to analytical precision; it should be concluded that the notion that manipulation negates voluntariness is too strong.

Agent-centeredness

I believe that grounding the concept of voluntariness *outside* of agents, as the VaC theory does, is inherently mistaken. The VaC theory essentially disregards the agent, as it is oriented around others' control of her rather than being oriented around her directly; it is not agent-centered. But, voluntariness is deeply connected to individual agents' engagement with their choices. Therefore, it is crucial that the agent's subjective point of view is not neglected in a conceptualization of voluntariness.

My proposal is that an action should be understood as non-voluntary if and only if it is coerced, and if the agent thinks of the action as non-voluntary when being coerced to perform it. In the following sections, I will attempt to establish the view that the agent's subjective point of view is important to a conceptualization of

voluntariness. Then, I will defend my proposal from three objections that I expect from critics.

Consider these examples, intended to demonstrate that the VaC theory disconnects the concept of voluntariness from agents' choosing and acting, and that it is problematic to do so.

Suppose (1) that a doctor informs her patient that she may choose between treatments *A*, *B*, and *C*. The hospital has financial interests in patients choosing between *A* and *B*, so the doctor has been instructed to manipulate the patient by purposefully depicting *C* as less favorable than it actually is and *A* and *B* as more favorable than they actually are; the doctor complies. Now, suppose (2) that the doctor does not purposefully depict *A*, *B*, and *C* as such, but that she mistakenly believes them to have the characteristics that she attributes to them in (1).

VaC theorists would argue that the patient is being controlled in (1) and therefore chooses non-voluntarily. However, in (2), the patient's decision is not being controlled and she is therefore choosing voluntarily. But, the patient made what she conceived of as two identical choices: She does not think of them as non-voluntary. Theories that give different answers to (1) and (2) neglect her point of view. They detach the concept of voluntariness from agents' choosing and acting.

Now, suppose (3) that the doctor tries to follow the hospital's instructions but becomes confused and, while believing that she is successfully deceiving the patient, she informs her about *A*, *B*, and *C* as they actually are. In (3), VaC theorists would again claim that the patient's choice is being controlled by the doctor, if poorly so. They would argue that the patient chose non-voluntarily, despite the fact that the doctor's intended manipulation had no effect on her choosing. The patient made the same choice as she would have made had she been treated by a non-manipulative doctor. VaC theorists, and others, must explain why a person's action *Y*, that has no effect on another person's action *Z*, renders *Z* non-voluntary disregarding of the *Z*-person's own beliefs. To be clear, the VaC theory has the absurd consequence that there could be situations when VaC theorists must try to convince people whose stable and considerate opinions are that they have acted voluntarily that they are mistaken: "Your views do not matter, you acted non-voluntarily."

An agent-centered theory could deal with these examples, as it would incorporate the agent's experience of choosing and acting. Therefore, the VaC theory is suboptimal in comparison to agent-centered theories in at least one respect.

My proposal

To repeat, I propose that an action should be understood as non-voluntary if and only if it is coerced, and if the agent thinks of the action as non-voluntary when being coerced to perform it. There are three main complexities with this conceptualization: It is not immediately clear what it is to think of an act as non-voluntary, why it should be thought when being coerced, or what function the notion of coercion has. They will be elaborated on in that order.

Not everyone has a stable and considered idea of what voluntariness is. Yet, my proposal seems to require just that, since the proposal is that agents must think of their actions as non-voluntary for them to be non-voluntary. What I have in mind though, is an agent's inner sense of resistance to the behavior that she is exposed to; an intolerance, anti-attitude, or unwillingness.

The inner sense must not be displayed by the agent, it suffices that she experiences it. An inner sense of resistance is displayed either verbally by articulating it, or physically by resisting. Neither should be required of agents for their actions to be labeled non-voluntary. Suppose, for instance, that a doctor points a gun to the head of a patient who refuses treatment with the intention of forcing her to undergo it. It is unreasonable to say that the patient then undergoes treatment voluntarily only because she did not protest or resist when the doctor pointed a gun at her. It is enough that the patient experiences a sense of inner resistance for the action to be non-voluntary, she must not also display it.[8]

The agent should think of the action as non-voluntary when being coerced to perform it. More precisely, it should be thought at least at some point while the coercion lasts. It could be the case that a person is coerced to undergo a long treatment during which her inner resistance slowly breaks down until she becomes emotionless or apathetic towards it. That does not mean that she continues the treatment voluntarily. But, if the person changes her attitude not due to the coercion, or to the influence of some other person or condition, but due to her own independent and rational revaluation, it is not unreasonable to say that she continues her treatment voluntarily (although she begun it non-voluntarily). Furthermore, if an agent thinks of an action as non-voluntary first after the coercion has ended, it was not non-voluntarily performed. A conceptualization of voluntariness should not enable agents to dress up

[8] It is mainly through the notion of thinking of one's own actions as non-voluntary that the agent's perspective is taken into account by my proposal, i.e., what makes my proposal agent-centered.

their regret in an exculpating language that downplays their own role in what has unfolded.

Finally, an action is non-voluntary only if it is coerced. Coercion is a form of influence not like any other. It alone bears the capacity to neutralize agency completely. Therefore, while other forms of influences can be gravely unjustified, they do not reach the same threshold as coercion does. Yet, I withhold that coercion does not negate voluntariness by default. This is unconventional and requires an explanation.

Coercion and voluntariness

Suppose that a patient faces the treatment options *A*, *B*, and *C*, that she ranks them in that order, but that she thinks of them as three good options, much unlike option *D*, *E*, and *F*, which she thinks of as bad options. Now, suppose also that the patient's doctor for some reason wants the patient to choose option *B*. Upon realizing that the patient is about to choose treatment *A*, the doctor bursts out: "If you choose *A*, I will reduce your dose of painkillers!" The patient, who fails to appreciate the peculiarity of the doctor's unprovoked outburst, shrugs disinterestedly. "Okay," she says, "I'll choose *B* then."

It is true that the patient chooses *B* non-autonomously, but although the doctor coerces her (per VaC standards), it is alien to how the patient leads her life to say that she chooses B non-voluntarily. Had the doctor forced the patient to choose between *D*, *E*, and *F*, it is likely that she would have thought of it as non-voluntary at the time of choosing, i.e., experienced an inner sense of resistance of the kind discussed above. But, since she thought of *A*, *B*, and *C* as good options, it did not dawn on her at the time that she might be wronged. If she later recollects the scene, she would be correct in saying that the doctor wronged her in some way, but incorrect in saying that she did not choose voluntarily; it would be an inaccurate description of something that should include her attitude toward the choice.

The example is intended to show that there is a difference between coercing someone who is engaged with a choice and coercing someone who is not. While it is possible, in most cases even likely, that a coerced agent is wronged in some way disregarding of whether she is engaged with her choice, the latter does not negate voluntariness. Whatever it is that makes it wrong to coerce people, the term "voluntariness" is not always the most suitable to employ in an articulation of the wrongness.

Objections

The first objection targets the subjectivism of my proposal, the second is that my proposal is counterintuitive, and the final objection is that my proposal is unconventional. I will answer to them in that order.

There are two aspects of the objection that targets the subjectivism of my proposal; one is that subjectivism entails relativism, the other is that subjective experiences cannot be objectively evaluated. Whether an action is voluntary, it might be argued, should be determined by objective standards. Resting an account on individual perceptions is untrustworthy and allows for relativism. This objection is mistaken. First, the notion of thinking about an action as non-voluntary is not intended to single-handedly explain voluntariness, as all actions thought of as non-voluntary by the acting agent would then also be non-voluntary. It is one of two necessary and jointly sufficient criteria that in conjunction explains voluntariness (coercion being the other). Second, the fact that my notion includes *subjective judgments* does not entail that it builds on *subjectivism*. I have proposed an objective standard for what it is to think of an action as non-voluntary, and agents can fail to adhere to it. Thereby, my notion is not inherently untrustworthy and does certainly not allow for relativism.

Regarding the objection that subjective experiences cannot be evaluated, my answer is that a distinction must be made between *criteria of evaluation* and *the decision procedures* by which such criteria are evaluated. The criterion of evaluation is here the notion that an action is non-voluntary only if the agent thinks of it as such when being coerced to perform it. It rests on whether an agent experiences an inner sense of resistance, which is epistemically difficult to determine. This epistemic difficulty is here a matter of decision procedure, i.e., of deciding whether the criterion of evaluation is met in a particular instance. But, decision procedures are unrelated to the truth-value of criteria of evaluation. The truth-value of the notion that an action is non-voluntary only if the agent thinks of it as such when being coerced does not hinge on whether the agent's thoughts are epistemically available to an observer. The fact that my account of voluntariness is accompanied by epistemic difficulties does not mean that it is false.[9]

The next objection is that my proposal is counterintuitive. In most theories of voluntariness, coercion negates voluntariness by default. Nonetheless, I think that

[9] And, the epistemic *difficulty* does not necessarily amount to an epistemic *impossibility*.

there are strong reasons to adhere to the notion that an action is non-voluntary only if the agent thinks of it as such when being coerced. Therefore, regarding coercion, I will have to bite the bullet. In the above section, I have presented an example showing that it is perhaps not completely unreasonable that an agent can be subject to coercion and still act voluntarily. It is a matter of being (or not being) engaged with the choice in question; if the choice does not matter to the agent, coercion does not affect the voluntariness with which she chooses. Even if my arguments are unconvincing, it should be concluded that the standard view that coercion negates voluntariness by default should not be uncritically accepted.

The final objection is that my proposal is unconventional. Voluntariness is one of bioethicists' favorite go-to-concepts. However, I think that it is over-used. Other VaC theorists than those already mentioned have suggested that potentially constraining political, cultural, social, and economic factors may limit an agent's degree of control over her choices and thereby the voluntariness with which she chooses (Bull and Lindegger 2011, p. 27). Yet others have expanded the theory further. For instance, it has been shown that in research involving children many mothers consider decision-making by the child's father as socially normative, and some argue that this norm influences the mothers' voluntariness (Molyneux et al. 2005; Kamuya et al. 2011, p. 31). Lastly, in perhaps the most demanding version of the VaC theory, it has been argued that voluntariness requires authenticity, understood as 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). In other words, bioethicists have attempted to squeeze almost anything into the concept of voluntariness, as if it could explain all sorts of moral dilemmas.

However, the concept is not as applicable as bioethicists often seem to think. On the contrary, most actions are voluntarily performed, while potential moral problems associated with them are better explained through other concepts. Bioethicists should strive toward conceptual diversity and plurality, developing more concepts with better precision, rather than attempting to fit everything into one universal tool.

One tool that might have to be invented is a criterion of engagement, or interest, in informed consent. That is, if voluntariness should be understood more narrowly than what is common, which is my view, something else might need to fill the newly formed space in between already established informed consent criteria. Perhaps people cannot consent to a medical intervention or research participation if they are not engaged in the decision, in the sense discussed above.

Putting the theory to work

The question remains how to put my theory to work. I will briefly demonstrate that here.

The patient in the coffee example takes her medicine voluntarily but unknowingly. As she is unaware of the medicine in her coffee she does not consent to being medicated.

Had the nurse said, "drink your coffee or I will not let your family in during visiting hours tomorrow," the patient would have been coerced. If the patient engages in the decision, as discussed above, and if the nurse had been successful in her effort to medicate the patient, it would have been an example of coercion that negates voluntariness. If the patient does not engage in the choice, but only shrugs disinterestedly, it would have been coercion that did not negate the voluntariness of her actions. Though, due to her ignorance of the medicine dissolved in the coffee, the patient's actions would not amount to valid consent in either case.

Other examples are: A patient who hears voices in her head telling her to consent to treatment consents voluntarily but incompetently; a patient who consents to treatment because "everyone expects her to" consents voluntarily but non-autonomously; a patient who refuses treatment because she wants to die does so voluntarily.

Concluding remarks

This article is an attempt to stimulate bioethicist debate on voluntariness; its conditions, and role. It has been suggested that voluntariness should cover lesser moral ground than what is conventional. Voluntariness should be an agent-centered concept that takes the agent's point of view into consideration through the notion that an act is non-voluntary only if the agent thinks of it as such when being coerced to perform it. Thereby, greater analytical precision can be achieved while voluntariness is reconnected with how people actually lead their lives, which fosters autonomy.

Bibliography

Ahlin, J. (forthcoming). "The impossibility of reliably determining the authenticity of desires: implications for informed consent." In: *Medicine, Health Care and Philosophy*. DOI: 10.1007/s11019-017-9783-0.

Appelbaum, P. S., C. W. Lidz, and R. Klitzman (2009). "Voluntariness of consent to research: A conceptual model." In: *Hastings Center Report* 39.1, pp. 30–39.

Archard, D. (2016). "The Methodology of Applied Philosophy." In: *A Companion to Applied Philosophy*. Ed. by K. Lippert-Rasmussen, K. Brownlee, and D. Coady. John Wiley & Sons, Ltd.

Aristotle. (2014). *Nicomachean ethics*. Cambridge University Press.

Banner, N. F. and G. Szmukler (2013). "'Radical Interpretation' and the Assessment of Decision-Making Capacity." In: *Journal of Applied Philosophy* 30.4, pp. 379–394.

Beauchamp, T. L. (2010). "Autonomy and Consent." In: *The ethics of consent: Theory and practice*. Ed. by F. G Miller and A Wertheimer. Oxford: Oxford University Press, pp. 55–79.

Beauchamp, T. L. and J. F. Childress (1983). *Principles of biomedical ethics*. 2nd ed. Oxford: Oxford University Press.

— (2013). *Principles of biomedical ethics*. 7th ed. Oxford: Oxford University Press.

Berghmans, R. (2011). "Voluntary consent, normativity, and authenticity." In: *American Journal of Bioethics* 11.8, pp. 23–24.

Björk, J., N. Lynøe, and N. Juth (2016). "Empirical and philosophical analysis of physicians' judgments of medical indications." In: *Clinical Ethics* 11.4, pp. 190–199.

Bowman, D., J. Spicer, and R. Iqbal (2012). *Informed consent: A primer for clinical practice*. Cambridge: Cambridge University Press.

Brennan, T. (1991). *Just doctoring: medical ethics in the liberal state*. University of California Press.

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

Bull, S. and G. M. Lindegger (2011). "Ensuring consent to research is voluntary: How far do we need to go?" In: *The American Journal of Bioethics* 11.8, pp. 27–29.

Burns, J. M. and R. H. Swerdlow (2003). "Right Orbitofrontal Tumor With Pedophilia Symptom and Constructional Apraxia Sign." In: *Archives of Neurology* 60.3, pp. 437–440.

Charland, L. C. (2001). "Mental competence and value: The problem of normativity in the assessment of decision-making capacity." In: *Psychiatry, Psychology and Law* 8.2, pp. 135–145.

— (2015). *Decision-Making Capacity*. URL: https://plato.stanford.edu/entries/decision-capacity/.

Christman, J. (2004). "Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves." In: *Philosophical Studies* 117.1, pp. 143–164.

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

— (2015). *Autonomy in moral and political philosophy*. URL: http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/.

Colburn, B. (2008). "Debate: The concept of voluntariness." In: *Journal of Political Philosophy* 16.1, pp. 101–111.

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*. URL: http://plato.stanford.edu/entries/informed-consent/.

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

Fisher, J. A. (2013). "Expanding the frame of 'voluntariness' in informed consent: Structural coercion and the power of social and economic context." In: *Kennedy Institute of Ethics Journal* 23.4, pp. 355–379.

Frankfurt, H. (1971). "Freedom of the Will and the Concept of a Person." In: *The Journal of Philosophy* 68.1, pp. 5–20.

Freedman, B. (1981). "Competence, Marginal and Otherwise: Concepts and Ethics." In: *International Journal of Law and Psychiatry* 4.1–2, pp. 53–72.

Gillon, R. (2003). "Ethics needs principles—four can encompass the rest—and respect for autonomy should be "first among equals"." In: *Journal of Medical Ethics* 29.5, pp. 307–312.

Grisso, T., P. S. Appelbaum, and C. Hill-Fotouhi (1997). "The MacCAT-T: A Clinical Tool to Assess Patients' Capacities to Make Treatment Decisions." In: *Psychiatric Services* 48.11, pp. 1415–1419.

Hippocrates. *Hippocratic oath*. URL: http://medical-dictionary.thefreedictionary.com/Hippocratic+oath.

Ho, A. (2008). "The Individualist Model of Autonomy and the Challenge of Disability." In: *Bioethical Inquiry* 5.2–3, pp. 193–207.

Hope, P. T. et al. (2011). "Anorexia Nervosa and the Language of Authenticity." In: *Hastings Center Report* 41.6, pp. 19–29.

Hyman, J. (2013). "Voluntariness and choice." In: *The Philosophical Quarterly* 63.253, pp. 683–708.

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.

Kamuya, D., V. Marsh, and S. Molyneux (2011). "What we learned about voluntariness and consent: Incorporating 'background situations' and understanding into analyses." In: *The American Journal of Bioethics* 11.8, pp. 31–33.

Kant, I. (1996). *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant)*. Cambridge University Press.

Katz, J. (1984). *The silent world of doctor and patient*. The Johns Hopkins University Press.

Loss aversion. URL: https://www.behavioraleconomics.com/mini-encyclopedia-of-be/loss-aversion/.

Lynøe, N. (1999). *Mellan cowboyetik och scoutmoral: Medicinsk forskningsetik i praktiken*. Liber.

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

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

Millenson, M. L. (2011). "Spock, feminists, and the fight for participatory medicine: a history." In: *Journal of Participatory Medicine* 3.27.

Miller, B. L. (1981). "Autonomy & the Refusal of Lifesaving Treatment." In: *The Hastings Center Report* 11.4, pp. 22–28.

Miller, V. A. et al. (2011). "The decision-making control instrument to assess voluntary consent." In: *Medical Decision-making* 31.5, pp. 730–741.

Molyneux, C. S. et al. (2005). "'Even if they ask you to stand by a tree all day, you will have to do it (laughter)…!': Community voices on the notion and practice of informed consent for biomedical research in developing countries." In: *Social Science and Medicine* 61.2, pp. 443–454.

Murray, P. M. (1990). "The History of Informed Consent." In: *The Iowa Orthopaedic Journal* 10, pp. 104–109.

Nelson, R. M. et al. (2011). "The concept of voluntary consent." In: *The American Journal of Bioethics* 11.8, pp. 6–16.

Nozick, R. (1969). "Coercion." In: *Philosophy, science, and method: Essays in honor of Ernest Nagel*. Ed. by W Morgenbesser. New York: St Martin's Press, pp. 440–472.

— (1974). *Anarchy, state, and utopia*. Blackwell Publishing.

Olsaretti, S. (1998). "Freedom, force and choice: Against the rights-based definition of voluntariness." In: *The Journal of Political Philosophy* 6.1, pp. 53–78.

— (2008). "Debate: The concept of voluntariness—A reply." In: *The Journal of Political Philosophy* 16.1, pp. 112–121.

O'Neill, O. (2003). "Some limits of informed consent." In: *Journal of Medical Ethics* 29, pp. 4–7.

Oshana, M. A. L. (2015). "Is Social-Relational Autonomy a Plausible Ideal?" In: *Personal Autonomy and Social Oppression: Philosophical Perspectives*. Ed. by M.A.L Oshana. 1st ed. Routledge, pp. 3–24.

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

Osuji, P. I. (forthcoming). "Relational autonomy in informed consent (RAIC) as an ethics of care approach to the concept of informed consent." In: *Medicine, Health Care and Philosophy*. DOI: 10.1007/s11019-017-9789-7.

Siedentop, L. (2014). *Inventing the Individual: The Origins of Western Liberalism*. Penguin Books.

Sjöstrand, M. and N. Juth (2014). "Authenticity and psychiatric disorder: does autonomy of personal preferences matter?" In: *Medicine, Health Care and Philosophy* 17.1, pp. 115–122.

Strong, C. (2013). "Medicine and Philosophy: The Coming Together of an Odd Couple." In: *The Development of Bioethics in the United States*. Ed. by J. R Garret, F Jotterand, and D Ralston. Springer Netherlands.

Sutrop, M. (2011). "Changing Ethical Frameworks: From Individual Rights to the Common Good?" In: *Cambridge Quarterly of Healthcare Ethics* 20.4, pp. 533–545.

Tan, D. J. O. A. et al. (2006a). "Competence to make treatment decisions in anorexia nervosa: thinking processes and values." In: *Philosophy, Psychiatry, & Psychology: PPP* 13.4, pp. 267–282.

— (2006b). *Competence to make treatment decisions in anorexia nervosa: thinking processes and values*. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2121578/pdf/nihms5623.pdf.

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

Taylor, J. S. (2009). *Practical Autonomy and Bioethics*. Routledge.

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

Wall, E. (2001). "Voluntary action." In: *Philosophia* 28.1–4, pp. 127–136.

Wertheimer, A. (2012). "Voluntary consent: Why a value-neutral concept won't work." In: *Journal of Medicine and Philosophy* 37.3, pp. 226–254.

Widdows, H. (2011). "Localized past, globalized future: towards an effective bioethical framework using examples from population genetics and medical tourism." In: *Bioethics* 25.2, pp. 83–91.

Wiking, E. (2012). *Health Care on Equal Terms for Immigrants in Sweden*. Karolinska Institutet.

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

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

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.