Distribution Agreement

Signature:

______________________________	____________________
Brian Scott Jenkins	Date

The Moral Significance of Non-Autonomous Refusals of Medical Treatment

By

Brian Scott Jenkins
Master of Arts

Bioethics

Toby Schonfeld, Ph.D.
Advisor

Cory Labrecque, Ph.D.
Committee Member

Mark Risjord, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

The Moral Significance of Non-Autonomous Refusals of Medical Treatment

By

Brian Scott Jenkins
B.S., Georgia Southern University, 2006

Advisor: Toby Schonfeld, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Bioethics
2015

Abstract

The Moral Significance of Non-Autonomous Refusals of Medical Treatment

By Brian Scott Jenkins

In contemporary medical ethics, the theoretical justification for informed consent has been extended to treatment refusal decisions. This has rendered a theoretical framework in which compulsory treatment of non-autonomous patients (those lacking decision-making capacity) is justified solely by reference to their lack of decision-making capacity. The result is that an active refusal by a non-autonomous patient may be automatically overridden by the consent of a surrogate decision maker. The primary objective of this thesis is to make the case that the nature of an active refusal is such that lack of decision-making capacity should not automatically negate one's right to refuse medical treatment. I will argue that a refusal contains a moral significance, the source of which is one's negative liberty rights, that is independent of a patient's decision-making capacity. Therefore, if a non-autonomous patient is to be treated against his or her expressed wishes, the treatment must be justified on grounds in addition to his or her lack of decision-making capacity. The secondary objective of this thesis is to propose a principle that balances the moral significance of a non-autonomous patient's refusal with the obligation to protect him from the consequences of his decisions. I will argue that a harm principle accomplishes this task and is the appropriate threshold for determining whether overriding the patient's expressed refusal is justified.

The Moral Significance of Non-Autonomous Refusals of Medical Treatment

By

Brian Scott Jenkins
B.S., Georgia Southern University, 2006

Advisor: Toby Schonfeld, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Bioethics
2015

Acknowledgments

I would first like to thank my advisor, Toby Schonfeld, for her guidance and support during this process. Her critical eye was an invaluable resource for turning what began as an intuitive discomfort into a fully-articulated thesis. I would especially like to thank her for continuing as my advisor after she left Emory. Despite her demanding schedule, she continued to provide helpful feedback as I sent her drafts. I would also like to thank my other two thesis committee members, Cory Labrecque and Mark Risjord, for their insight and feedback. My bioethics classmates deserve a special thanks for their encouragement as we progressed through the program together. In particular, I would like to thank Sarah Putney, my former boss and classmate, who shared her experiences as we were concurrently writing our theses. I would like to thank my family and friends for supporting me through all of my educational endeavors. Finally, I'd like to thank my girlfriend, Martha, for her support and understanding when my non-work hours were spent writing my thesis.

Table of Contents

CHAPTER 1

Consider the case of Ryan, a twenty one-year old college student who lives with his roommate, Jack. Prior to moving in together six months ago, Ryan was a normal college student who made good grades and enjoyed a vibrant social life. Lately, he has become more isolated, locking himself in his room and foregoing activities that he normally enjoyed. Sometimes late at night, Jack hears Ryan walking around the apartment and moving the blinds to look out the windows. Ryan has also shown mild signs of suspiciousness and hostility toward Jack, but he has not exhibited any violent tendencies. Jack spoke to his mother, a psychiatric nurse, about the situation, and she told him that it sounds like the early signs of schizophrenia and that Ryan needs to be seen by a psychiatrist. Jack repeatedly attempted to persuade Ryan to see a doctor, but the attempts always failed. One night when Jack returned home from dinner with his girlfriend, Ryan had all the lights out as he sat in the living room looking out the window. He was visibly distraught, his speech was only semi-coherent, and he told Jack that the government was controlling his mind with satellite signals. He also believed that government agents were outside planning a raid on their apartment. Not knowing what to do, Jack called 911 and requested help. The EMTs brought Ryan to the nearest hospital where he was placed in the psychiatric unit. An assessment by the attending psychiatrist revealed that Ryan was suffering from an acute schizophrenic episode and that he should be treated with antipsychotic medication. Ryan had no insight into his disease. He believed the psychiatrist was a CIA agent who was trying to kill him because he had damaging information against the government. He did not consent to being treated with the antipsychotics because he thought they were intended to kill him. Despite his explicit refusal, Ryan was treated against his wishes because he was judged not to have the capacity to make his own medical decisions.

Informed Consent, Decision-Making Capacity, and Refusal

Since Ryan was judged to lack decision-making capacity, the decision to treat him against his expressed refusal is not likely to be judged as a moral violation by contemporary ethical standards. In medical ethics, a patient's autonomy is respected through the requirement that informed consent be obtained before proceeding with treatment. Several conditions are necessary for informed consent to be morally valid. Faden and Beauchamp argue that informed consent "is given if a patient or subject with (1) substantial understanding and (2) in substantial absence of control by others (3) intentionally (4) authorizes a professional (to do Intervention I)" (Faden and Beauchamp 1986, 278). Meisel and colleagues provide similar criteria: the patient's decision must be voluntary; the patient must be provided sufficient information to make a decision; the patient must be competent and understand the information provided; and the patient must, ultimately, consent (Meisel, Roth and Lidz 1977).

Variations of these conditions have been used in medical ethics to solicit what has come to be called "informed refusal." Intended to ensure that a patient's decision to forgo medical treatment is autonomous, such conditions impose the same criteria of information disclosure, competence, patient comprehension, and voluntariness that are fundamental to the concept of informed consent (The American College of Obstetricians and Gynecologists 2004). Refusal and consent are, thus, conceived as symmetrical aspects of respect for patient autonomy. Under such a construal, they are treated as essentially two sides of the same coin, where the criteria for a respect-worthy refusal are essentially the same as those for respect-worthy informed consent. They have been united conceptually under the notion of medical decision-making, where refusal is seen more as *consent to no treatment* than simply a refusal to consent to (or permit) the treatment. Of course, a patient who refuses to consent to a specific treatment could consent to an alternative treatment, but as in Ryan's case, the issue is whether the patient may refuse to consent to *any* treatment. This sort of scenario, in which the patient decides to forego all treatment, is what characterizes the notion of an "informed refusal," where the patient is expected to understand the implications of opting for no treatment at all. Under such an expectation, the rubric of "decision-making capacity" has become a proxy for autonomous choice (Appelbaum, Assessment of Patient's Competence to Consent to Treatment 2007), whether that choice is to consent to treatment or to refuse to consent to treatment, and the right not to be treated against one's expressed refusal has become contingent upon the ability to rationally defend one's preference for being left alone. Thus, the basis for treating Ryan despite his expressed refusal was his inability to exhibit decision-making capacity by providing an acceptable justification for foregoing treatment. Because his physician determined that he lacked decision-making capacity, Ryan's expressed refusal of treatment was considered not to be autonomous, and his wishes, therefore, did not need to be respected.

Beauchamp and Childress offer two threshold elements, or preconditions, of informed consent, and, thus, "informed refusal," that are helpful for understanding the concepts that underlie autonomy: competence to understand and decide (decision-making capacity) and voluntariness in decision-making (Beauchamp and Childress 2009, 120). Simply put, competence requires the patient to have the capacity to understand a procedure, deliberate about its risks and benefits, and make a decision based on the deliberation (Beauchamp and Childress 2009, 114). A decision is made voluntarily when a person wills an action “without being under the control of another’s influence” (Beauchamp and Childress 2009, 132). According to Beauchamp and Childress, these two elements are common to virtually all theories of autonomy, though there is some variation in how the meanings of the two terms are understood. The two threshold elements described above represent two conditions essential for autonomy more generally, not just as it is implemented in medical ethics: *liberty* (independence from controlling influences) and *agency* (capacity for intentional action) (Beauchamp and Childress 2009, 100).

The two-pronged approach to autonomy proposed by Beauchamp and Childress provides a useful paradigm for understanding the relationship between decision-making capacity and informed consent (or refusal) and autonomy more generally. In the medical context, a patient must possess the capacity to make his or her own medical decision (the first precondition of informed consent/refusal), but the decision, itself, must be made voluntarily (the second precondition for informed consent/refusal). For the decision to be considered voluntary according to the standards of contemporary medical ethics, however, the patient must also be provided sufficient information about the treatment (Faden and Beauchamp 2012, 155). Otherwise, the patient would lack the information necessary to deliberate rationally about whether or not he or she should agree to the proposed treatment. With false or insufficient information about the treatment, the patient would not be in a position to understand the actual treatment, rationally deliberate about its actual risks and benefits, and make a decision based on that deliberation. Thus, his or her capacity to understand, deliberate, and make a decision would be irrelevant because the informational requirements would not have been satisfied, and any decision he or she made would not be considered autonomous because the decision would have been uninformed.

As Beauchamp and Childress note, respect for autonomous agents requires more than simply refraining from interfering in others' affairs, a *negative obligation* (Beauchamp and Childress 2009, 104). Respecting others as autonomous agents is to "acknowledge their right to hold views, to make choices, and to take actions based on their personal values and beliefs" (Beauchamp and Childress 2009, 103). It requires "building up or maintaining others' capacities for autonomous choice while helping to allay fears and other conditions that destroy or disrupt autonomous action" (Beauchamp and Childress 2009, 103). Thus, the informational requirements of informed consent are based on a view of respect for autonomy as medical professionals' *positive obligation* to patients (Beauchamp and Childress 2009, 104). In order for medical professionals to respect patients as autonomous agents, they must facilitate autonomous decision-making by ensuring that patients receive sufficient material information about the treatment, understand the information, voluntarily decide upon it, and, ultimately, authorize the medical professional to perform, or not perform, the treatment (Beauchamp and Childress 2009, 120-121). Patients who enjoy decision-making capacity, then, are autonomous agents when they enter the medical context, but they are unable to make an autonomous medical decision until the medical professionals have fulfilled their obligations to inform them about the proposed treatment.

Of course, the obligations of medical professionals to respect patients as autonomous agents are only present when the patients possess decision-making capacity. If a patient lacks decision-making capacity, he or she is not able to understand a treatment, deliberate about its risks and benefits, and make a decision. Accordingly, it is nonsensical to obligate medical professionals to ensure that patients perform actions for which they do not have the capacity. Further, contemporary medical ethics regards the negative obligation of medical professionals, to refrain from interfering in the affairs of others, to be absent as well when patients lack decision-making capacity. In the medical setting, when a patient lacks decision-making capacity, a “surrogate” decision-maker is appointed to make decisions on his or her behalf. The surrogate may be appointed by the patient or he or she could be selected by the court or another appropriate body (Berg, et al. 2001, 109-110). Once a surrogate decision maker is appointed, the contemporaneously expressed wishes of the patient are, at least in terms of the ethical justification for surrogate decision-making, inconsequential.

There are several suggested standards to guide surrogate decision-making. Beauchamp and Childress offer us three. The first is the standard of “substituted judgment.” According to this standard, a patient has the right to make decisions and to have his values respected, but he lacks the capacity to exercise this right. A surrogate would, therefore, decide on the basis of what the incapacitated patient would have wanted had he had decision-making capacity. This standard is based on the premise that the patient has the right to decide and have his values and preferences taken seriously, but he cannot exercise these rights because he is no longer, or has never been, competent. Beauchamp and Childress argue that the premise of the substituted judgment standard rests on a fiction, for an incompetent person cannot literally have the right to make medical decisions when a competent surrogate must exercise that right on his behalf. Further, they argue that the standard should be rejected for never-competent patients because no basis exists for judging what a non-autonomous person would have autonomously decided because he has never been autonomous. The substituted judgment standard, then, collapses into what Beauchamp and Childress call the "pure autonomy" standard, which applies exclusively to patients who once had decision-making capacity. According to this standard, the surrogate decides based on treatment preferences the once autonomous, but now non-autonomous, patient expressed in the past. Finally, there is the "best interests" standard. This standard obliges the surrogate to essentially decide based on a calculation of what will produce the greatest benefit to the patient. To the extent they are relevant to deciding what constitutes a *benefit*, the patient's preferences and values may play a role in the surrogate's decision (Beauchamp and Childress 2009, 135-139). Regardless of which standard is used, however, the expressed wishes of the patient[1] at the time of the medical interaction are superseded by the surrogate's decision.

The Structure of Autonomy

The standard practice in medicine with regard to patient decision-making, as described above, suggests a particular structure of autonomy. If the patient lacks *agency* in the form of medical decision-making capacity, then his right to *liberty*, or his right to act without the controlling influences of others, is also forfeited, and a surrogate decision maker will be appointed to decide on his behalf what will happen to him, regardless of

---

[1] If the patient is unable to express any wishes at the time of the medical interaction (e.g., if he is unconscious), then the surrogate's decision does not supersede his expressed wishes (because he has not expressed any). This thesis does not concern such cases, but rather only those in which the patient lacks decision-making capacity but is nonetheless able to express wishes regarding treatment.

his expressed wishes at the time. Thus, not only are *agency* and *liberty* necessary conditions for autonomy, but *agency* is a necessary condition for the exercise of *liberty*. As such, when a patient without capacity expressly refuses medical treatment, the primary justification for treating him against his expressed refusal is his lack of *agency*. Since agency is a necessary condition for autonomy, it is correct to conclude that the patient is non-autonomous, and that treating him against his expressed wishes is not an act which fails to respect his autonomy (because he is not autonomous).

Decision-Making Capacity

To possess *medical agency* is to possess medical decision-making capacity. To lack decision-making capacity is to lack *medical agency* and, thus, the right to refuse ostensibly medically-warranted treatment (*liberty*). Appelbaum and Grisso note that the legal standards for competence (or decision-making capacity) vary across jurisdictions, but they can generally fall into one or more of the following categories: the capacity to communicate a choice; the capacity to understand relevant information; the capacity to appreciate the situation and its consequences; and the capacity to manipulate information rationally (Appelbaum and Grisso, Assessing Patients' Capacities to Consent to Treatment 1988).

As they point out elsewhere, the use of a particular standard, or several standards simultaneously, could potentially have a significant effect on a determination of whether a patient is judged to have decision-making capacity. For instance, if a patient is deficient in the capabilities associated with all of the standards, then the use of any particular standard will have no effect on a capacity determination because the patient will lack decision-making capacity according to all of the standards. If a patient is deficient in the capabilities associated with one or more standards but not all, however, then the choice of standard(s) impacts whether that patient is determined to possess decision-making capacity (Grisso and Appelbaum, Comparison of Standards for Assessing Patients' Capacities to Make Treatment Decisions 1995). For instance, a patient may be capable of communicating a decision, or even understanding the information presented to him by the medical professional, but he may not be able to appreciate the situation and its consequences or manipulate the information rationally. If, say, understanding relevant information was the requisite standard, then the patient would seem to possess decision-making capacity. If the requisite standard was the ability to appreciate one’s own situation and the consequences of the decision, however, the patient would not possess the capacity to make his own decisions. Regardless of the legal standard, however, all of the capabilities are required for a person to give an informed consent and, thus, an informed refusal.

Relationship Between Agency and Liberty

Though the concepts of *agency* and *liberty*, as provided by Beauchamp and Childress, will be considered in more detail later, a preliminary reflection upon the relationship between autonomy in the medical sphere and autonomy more generally is perhaps necessary. A patient such as Ryan, who lacks the capacity to make a decision regarding medical treatment of his schizophrenia episode, is not capable according to contemporary standards of medical ethics of refusing treatment. As Beauchamp and Childress point out, decision-making capacity[2] is task- or decision-specific, and the criteria for determining whether one has the capacity (or competence) varies from context

---

[2] Beauchamp and Childress actually use the term *competence* here because, they argue, the distinction between *competence*, a legal status, and *capacity*, a patient’s ability as judged by a medical professional, breaks down in practice.

to context (Beauchamp and Childress 2009, 112). Ryan, for example, may be perfectly capable of deciding what to wear or what he wants for dinner, but he still does not possess sufficient decision-making capacity to refuse, or even opt for, treatment of his illness. This is also evident in the legal context, where courts generally make task-specific determinations of competence or incompetence (Leo 1999). For instance, an individual may be determined competent to stand trial but determined incompetent to make his or her own medical decisions.

The rationale for using a task- or decision-specific approach to decision-making capacity is to recognize that the patient may be capable of making many different types of decisions even if he or she is incapable of making a medical decision. To avoid generalizing the findings of a capacity determination to other spheres of decision making, then, it is necessary to restrict the focus to the particular decision(s) under scrutiny (Beauchamp and Childress 2009, 112). Since individuals are presumed to have agency until proven otherwise, this has the purported purpose of respecting even patients who have been judged to lack medical decision-making capacity as autonomous agents, at least to the extent they are capable of making non-medical decisions (or even less crucial medical decisions for which they may have capacity). Patients, then, should be free to pursue courses of action that fall outside of the domain of the particular medical decision with which they are confronted.

In practice, however, the specificity of a decision-making capacity determination does not necessarily constrain the limits placed on an incapacitated patient's otherwise autonomous decisions. Rather, it reflects a presumption of the primacy of the medical decision over all other realms of decision making and action. According to the structure of autonomy described above, a patient who lacks decision-making capacity (agency) does not enjoy the right to refuse medical treatment for his or her illness (liberty). Since medical decision-making capacity is task- or decision-specific, the patient's *medical agency* is a very narrow notion, confined to a specific decision or limited set of decisions related to his or her medical care. Accordingly, the corresponding *medical liberties*, those which relate directly to the specific medical decision, are also narrow. If the patient lacks medical decision-making capacity, he or she does not enjoy liberty with regard to any medical care about which he or she is not capable of making an informed decision (consent or refusal). Despite being so narrow in content, medical autonomy operates as a gatekeeper to the enjoyment of autonomy more generally once one finds oneself in the medical context. If you assume, for instance, that Ryan is autonomous in every respect except with regard to making a decision about treatment for his schizophrenia, then it becomes clear that there are spheres of decision making and action in which he is not permitted to participate even though he is capable of doing so. If Ryan wanted to, say, move to a commune in New England or exercise his right to vote, he would be unable to do so because he lacks the *medical agency* necessary to disengage from the medical context. The contemporary justification for compulsory treatment, then, presupposes a hierarchy in which autonomous medical decisions are of primary importance, and other spheres of decision making and action are relegated to a more subordinate class.

The potential for differences in a patient's capacity to make different decisions, particularly medical versus non-medical decisions, could stem from a variety of sources. There are at least three distinct, albeit interrelated, possible sources of such differences. The first could be that there are inherent differences between certain types of decisions, and the functional capabilities required for sufficient decision-making capacity vary with the types of decisions. In other words, the nature of the particular decision faced by an individual (with a specific set of functional capabilities) has an impact on whether or not he or she possesses agency and, therefore, is able to act autonomously.

The second possible source is related to the first. If there are inherent differences between particular types of decisions, and those differences are the source of distinctive requirements for agentive action, then there should be different standards for determining whether one possesses agency with respect to a particular decision or action. In practice, there are differing standards for different types of decisions. In the medical context, patients must be capable of making an “informed” decision based on relevant information. The standards for deciding whether one has the capacity to decide whether to buy a vehicle or make a particular investment, however, are not quite so stringent. Though there may be certain disclosure requirements incumbent upon the salesman or financial advisor, there is no obligation to determine whether the person has the capacity to make such a decision based on the information disclosed. Moreover, there is certainly no requirement that the person demonstrate the capacity to refuse to engage in such a transaction, even if there are compelling reasons for him to do so. Of course, the fact that the standards do vary with respect to the particular type of decision does not necessarily mean that the nature of the particular decision warrants its corresponding standard of scrutiny. Medical decisions can be very complex and their consequences can have a significant impact on the patient’s wellbeing, so it seems justifiable to require medical professionals to disclose relevant information and ensure that a patient’s decision is made voluntarily and through a sound decision-making process. It could be argued, however, that some non-medical decisions such as a vehicle purchase or investment decision, or even voting, are sufficiently complex and have significant enough consequences that they warrant decision-making capacity standards similar to those necessary to demonstrate *medical agency*. Nonetheless, the standards for demonstrating agentive capacity with regard to these decisions are minimal if not non-existent.

The third possible source concerns the liberty to make certain decisions more so than it does an individual's decision-making capacity, or agency. If certain non-medical decisions actually are sufficiently similar to medical decisions, it should not seem unreasonable to impose similar decision-making capacity (agency) requirements on such decisions. Further, it would not seem a violation of one's right to autonomy if he were not allowed to act on his expressed wishes with regard to these decisions. Yet, these non-medical decisions are not governed by the same decision-making capacity (agency) criteria as medical decisions. As such, individuals lacking agency, at least with respect to the standards of contemporary medical ethics, are free to act on their non-medical wishes without external interference. In other words, even if there are non-medical decisions of the same nature as medical decisions, there are no apparent non-medical examples in which individuals are both held to medical ethics' stringent agency requirements and deprived of their liberty to act on their wishes based on their inability to satisfy those requirements.

The potential for various sources affecting an individual's agency with respect to a particular decision or course of action means that the normative requirements for agency and, thus, autonomy, are highly contextual in practice, and it is not immediately apparent which source is driving the contextual variation. It could be that the decision-making capacity criteria of contemporary medical ethics are the “gold standard” of agency and that all decisions should be judged according to such a standard. On the other hand, it could be that the standards for agency with respect to certain decisions flow logically from the nature of the decisions themselves, and that those who endorse or hinder specific decisions do so with a legitimate moral authority derived from the standards. If the first statement is correct, then any decision-making process that falls short of the standard set forth by contemporary medical ethics reflects a lack of agency, whether permanent or temporary, on the part of the decision maker. This would mean that other, non-medical spheres of decision-making, such as purchasing, investment, or voting decisions, could justifiably be subject to the jurisdiction of certain persons or entities with a legitimate moral authority to enforce the “gold standard.” If the second statement is correct, then the different standards for judging agency that we observe in practice could, at least theoretically, be justified. Assuming the standards of agency appropriately correspond to the nature of the particular decisions, the presence or absence of a certain authority to enforce a particular standard is morally defensible. It could be, however, that maintaining different standards for agency with regard to different types of decisions is morally appropriate but that the standards we observe in practice, and the presence or absence of certain authorities to enforce those standards, do not always appropriately correspond to the nature of the decisions themselves. If such were the case, then the instances in which the standards for agency do not match the nature of the decisions, resulting in an inappropriate presence or absence of certain authorities to enforce those standards, stand on an unstable moral foundation.

The Five Dimensions of Autonomy

As the discussion above suggests, the present notion of autonomy is dynamic in that it lies at the intersection of several dimensions, all of which influence whether a particular person facing a particular decision in a particular context is able to act autonomously. All of the dimensions are related to either an individual's agentive capacity with regard to a particular decision or course of action or to the capability and moral authority of certain individuals or entities to obstruct the expressed wishes of an individual who lacks such capacity. There are at least five such dimensions to consider, and they are in line with Beauchamp and Childress's two conditions for autonomy: agency (capacity for intentional action) and liberty (absence from controlling influences) (Beauchamp and Childress 2009, 100). The dimensions will be helpful for dissecting the concepts of agency and liberty and, hence, analyzing the case of Ryan, our example case of an individual suffering from an acute schizophrenic episode who has been determined to lack the requisite capacity to refuse treatment for his disorder.

The first dimension involves the nature of the decision at hand and whether the standard for agentive capacity appropriately corresponds to the nature of the decision. In medical ethics, a patient must be able to make an "informed" medical decision. Since the standards for morally-valid consent and refusal are the same, the nature of a refusal to submit to an unwanted treatment is considered equivalent to the nature of an agreement to undergo a particular treatment. As such, the standard for agentive capacity with regard to each decision is the same.

The second dimension involves whether the individual who faces the decision is capable of satisfying the criteria set forth in the prevailing standard for agentive capacity, assuming the standard appropriately corresponds to the nature of the decision. This dimension essentially concerns whether or not the individual possesses agentive capacity and is, thus, capable of making a free and self-directed decision. This is generally where analyses of treatment refusals begin in contemporary medical ethics (The American College of Obstetricians and Gynecologists 2004). It is assumed that the standard for the capacity to refuse medical treatment appropriately corresponds to the nature of the decision, and the discussion revolves primarily around whether or not an individual enjoys the capacity. Of course, it is not unreasonable to rely on such precedent, given the immediacy of the ethical issues encountered in everyday medical practice.

The third dimension includes any conditions or requirements external to the individual that are necessary for him or her to make an autonomous decision. For instance, in contemporary medical ethics, there is thought to be a *positive obligation* on the part of medical professionals to provide sufficient information about the treatment (or refusal) and to ensure that the patient understands it (Beauchamp and Childress 2009, 104-109). This could be thought of as a sort of facilitative dimension. In order for the individual to decide and act autonomously, certain conditions may be required, and there may be obligations on the part of other persons or entities to ensure those conditions are met. Whereas it is incumbent upon the medical professional to provide unbiased and sufficient information and to make sure the patient understands it, these conditions are not necessarily required in other contexts. A car salesman, for example, is required to disclose certain information, such as the price of the vehicle, but he is not required to do much else. Of course, he is morally, and perhaps legally, obligated to provide truthful information, but his job is to persuade you to buy a vehicle, not to facilitate an autonomous decision on par with that required in the medical realm.

Whereas the three dimensions above relate to an individual's agency and, hence, capacity to act autonomously, the fourth dimension, and the one that follows, concern the individual's liberty or freedom to act on his or her expressed wishes. The fourth dimension involves the presence or absence of an individual or entity with the capability of obstructing an individual's proposed course of action. Note that this dimension concerns only the presence or absence of an individual or entity, not whether or not the individual or entity justifiably prohibits an individual from acting on his or her expressed wishes. While a patient like Ryan is hospitalized, for example, medical personnel are present and they are capable of treating him against his wishes, or obstructing any other of his proposed courses of action, whether or not there is a satisfactory moral justification for doing so. They simply have the proximity to Ryan and infrastructure necessary to prevent him from acting on his wishes. On the other hand, if Ryan is home alone in his living room, there is no individual or entity with such capabilities present. He can do whatever he wants. He can commit suicide, mutilate his own body, or simply sit there and do nothing.

Finally, the fifth dimension concerns whether or not the individual or entity, if present, possesses the moral authority to obstruct an individual's proposed course of action, the source(s) of this authority, and its limits. The source of the individual or entity's moral authority to obstruct an individual's actions is not necessarily related to that individual's agentive capacities, even though this appears to be the case in the medical realm. The primary justification for treating Ryan against his expressed refusal is his apparent lack of agency, but in other contexts, notably those in which an individual possesses agency, obstructing that individual's proposed course of action must be based on another justification. Law enforcement officers, for example, possess the legal authority, which is presumably based upon a broader moral authority, to prohibit and/or prevent citizens from engaging in criminal behavior, whether or not the behavior is performed by someone who possesses agency. Of course, since this authority is limited to the enforcement of laws, law enforcement officers may not justifiably prevent or deter individuals from engaging in legal behavior, regardless of whether the behavior is immoral, in bad taste, or appears to be careless or imprudent.

Structure of Autonomy Along the Five Dimensions

Though it may be evident how the five dimensions fit into the structure of medical autonomy, it is especially important to recognize that, in contemporary medical ethics, the moral authority of medical professionals to force treatment on non-autonomous patients is based on certain factors pertaining to the second dimension as well as an acceptance of the notion that agency is a necessary condition for liberty. Put another way, the fifth dimension depends solely on the second dimension. In the next chapter of this thesis, I will argue this should not be the case. Instead, I will contend, the nature of an active treatment refusal is such that a lack of decision-making capacity should not automatically negate one's right to refuse medical treatment. By exploring the relationship between a consent and a refusal, and the moral authority of a surrogate to consent on a non-autonomous patient's behalf, I will argue that a refusal contains a moral significance that is independent of a patient's decision-making capacity.

In the third chapter of this thesis, I will attempt to locate the moral significance of a non-autonomous refusal in one's negative liberty rights. I will begin by using the work of Isaiah Berlin, particularly his influential essay *Two Concepts of Liberty*, to frame the current structure of autonomy in light of the notions of *positive* and *negative* liberty. I will then argue that Richard Fallon's notion of *ascriptive autonomy*, "the autonomy we ascribe to ourselves and others as the foundation of a right to make self-regarding decisions" (Fallon Jr. 1994, 878), signifies the moral significance of a non-autonomous refusal and underlies one's negative liberty rights. If non-autonomous refusals are morally significant and the moral significance is relevant to one's negative liberty rights, I conclude, then medical agency (or decision-making capacity) cannot be a necessary condition for the right to (negative) liberty. Decision-making capacity, therefore, should not serve as a comprehensive gatekeeper to one's right to refuse medical treatment.

Can Forced Treatment Be Justified?

The critique of compulsory treatment previewed above may appear to preclude a moral justification for treating a patient like Ryan against his expressed refusal. In the final chapter of this thesis, I will argue that though it questions the justification for forcibly treating a patient like Ryan, the argument against the contemporary structure of autonomy does not imply that there cannot be a moral imperative to treat a patient without decision-making capacity against his wishes. If there is indeed a moral imperative to treat a patient lacking decision-making capacity against his wishes, however, it must be justified on grounds in addition to his lack of agentive capacity. I will propose the use of a modified *harm principle* similar to the one suggested by Diekema in his article about state intervention in cases of parental refusal of medical treatment (D.

Diekema 2004). This would shift the burden to the medical professional to demonstrate not only that the patient lacks decision-making capacity but that his expressed desires are likely to lead to harm. The source of his moral authority, therefore, would not be solely dependent upon the patient’s lack of agency.

CHAPTER 2

Whereas the contemporary formulation of medical decision making presupposes a functional symmetry between consent and refusal in terms of agentive capacity requirements, in this chapter, I will argue that the nature of an active treatment refusal is such that a lack of capacity does not automatically negate one’s right to refuse medical treatment. By exploring the relationship between a consent and a refusal, and the moral authority of a surrogate to consent on a non-autonomous patient’s behalf, I will contend that a refusal contains a moral significance independent of a patient’s decision-making capacity. I will do so, in part, by differentiating an expressed (or active) refusal by a non-autonomous patient from cases in which a patient wishes to undergo treatment but lacks the capacity to consent.

Autonomy and Informed Consent

The primary justification for informed consent in medical ethics is respect for individuals’ autonomy (Beauchamp and Childress 2009, 118). Some theorists, such as Gerald Dworkin, argue in favor of an authenticity model of autonomy that focuses on autonomous *persons* (Dworkin 1988, 19-20). This framework views autonomy as a global rather than a local concept, meaning that it can only be assessed over extended portions of a person’s life (Dworkin 1988, 16). Since autonomous persons could fail to act autonomously in certain situations, Faden and Beauchamp argue that the best characterization of autonomy as it relates to informed consent is one that focuses on autonomous *actions* rather than autonomous *persons*. It is always an open question, they argue, whether an autonomous person, in any particular instance, has given informed consent by autonomously authorizing a medical procedure. A person who has the more broadly-conceived capacity to act autonomously could, for instance, fail to make an autonomous choice in a clinical setting if she is ill, manipulated, deceived, ignorant, overwhelmed by information, or for many other reasons, even if she signs an informed consent form (Faden and Beauchamp 1986, 237). I find Faden and Beauchamp's emphasis on autonomous *action* to be more suitable for analyzing autonomous informed consent. The provision of informed consent is a discrete *action*, so it is more fruitful to ensure that the analytical framework surrounding it corresponds to the parameters of that type of action. An analytical framework emphasizing a trait that endures, but to varying degrees, over time does not seem to do this. Therefore, I will use Faden and Beauchamp's approach to demonstrate the theoretical underpinnings of informed consent.

According to Faden and Beauchamp, there are three necessary conditions for autonomous action. A person acts autonomously only if that person acts: 1) intentionally, 2) with understanding, and 3) without controlling influences. Since acts are either intentional or unintentional, the condition of intentionality is not matter of degree. Therefore, whether or not an action is intentional has a direct effect on whether it is autonomous. If an action is intentional, it is potentially autonomous, but if it is not intentional, it is not autonomous. Understanding and non-control can be satisfied to a greater or lesser extent, however, so actions can be autonomous to a greater or lesser extent, depending on the degree to which these conditions are satisfied (Faden and Beauchamp 1986, 238). These three conditions can be understood in terms of Beauchamp and Childress's description of the components of theories of autonomy as follows: intentionality and understanding represent *agency* and non-control represents *liberty*.

Intentional action is "action *willed in accordance with a plan*, whether the act is wanted or not" (Faden and Beauchamp 1986, 243). A plan for action is a mental blueprint of the strategies proposed for the performance of the action. An intentional act must correspond to the actor's action plan, although the anticipated outcome might not materialize[3] (Faden and Beauchamp 2012, 242). An actor's intention in doing a certain act, therefore, must include a conception of how the particular act is to be done: "Whether a given act, X, is intentional, depends on whether in performing X the actor could, upon reflection, say, 'I did X as I planned,' and in that sense, 'I did the X I intended to do'" (Faden and Beauchamp 1986, 243).

In informed consent settings, the typical pattern for understanding is for patients to come to understand that they must consent to or refuse a proposed action by understanding what is communicated in an informational exchange with a medical professional (Faden and Beauchamp 1986, 250). Since the condition of understanding can be satisfied to varying degrees, full or complete understanding would be a polar extreme on the continuum. According to Faden and Beauchamp, an individual "has *full* or *complete* understanding of an action if there is a fully *adequate* apprehension of all the *relevant* propositions or statements (those that contribute in any way to obtaining an appreciation of the situation) that correctly describe: (1) the nature of the action, and (2) the foreseeable consequences and possible outcomes that might follow as a result of performing or not performing the action" (Faden and Beauchamp 1986, 252). Beauchamp

---

[3] Faden and Beauchamp do not elaborate on what they mean by this. However, my interpretation is that although intentional action in pursuit of an outcome does not guarantee that the outcome will happen, this does not mean that the action was not intentional. For instance, if someone intends to build a house, has a mental action plan for building a house, and executes that action plan only to produce something that does not resemble a house, he still intended to build a house. His action was intentional, but it did not produce the intended outcome.

and Childress's analysis of understanding echoes this view. They argue that "persons understand if they have acquired pertinent information and have relevant beliefs about the nature and consequences of their actions. Such understanding need not be complete, because a grasp of central facts is generally sufficient" (Beauchamp and Childress 2009, 127).

While the first two conditions present as positive occurrences, the condition of non-control is a negative condition that entails the absence of external controls on action. Faden and Beauchamp use the concepts of *influence* and *control* in their analysis of non-control. Control, they argue, is exerted through influence. Some influences are resistible, however, so the mere presence of an influence does not necessarily affect whether an action is autonomous. The polar extreme of a fully non-controlled act is an act that either has "(1) not been the target of an influence attempt, or (2) [if it has] been the target of an influence attempt, it was either not successful or it did not deprive the actor in any way of willing what he or she wishes to do or to believe" (Faden and Beauchamp 1986, 258). By contrast, a completely controlled act is wholly controlled by the will of another: "Person A's action controls an action X of person B if A gets B to do X through irresistible influences that would work even if B, left to his or her own ends, in no way wanted to do X" (Faden and Beauchamp 1986, 258).

There are three main categories of influence. Not all of them are controlling, and the controlling influences are not equally controlling. *Coercion* is a completely controlling form of influence. It occurs when "one party intentionally and successfully influences another by presenting a credible threat of unwanted and avoidable harm so severe that the person is unable to resist acting to avoid it" (Faden and Beauchamp 1986, 261). For an action to be coercive, it is crucial that the agent of influence *intend* to influence by presenting a threat to the other party. A mere perception of coercion in the mind of the other party is not sufficient to establish that coercion has taken place (Faden and Beauchamp 1986, 339). *Manipulation* is a wide-ranging category that includes intentionally and successfully influencing a person by non-coercively altering the actual choices available to him or her or by non-persuasively altering the person's perceptions of the available choices (Faden and Beauchamp 1986, 261). *Persuasion* is "the intentional and successful attempt to induce a person, through appeals to reason, to freely accept – as his or her own –the beliefs, attitudes, values, intentions, or actions advocated by the persuader" (Faden and Beauchamp 1986, 261-262). In contrast to the other two forms of influence, Faden and Beauchamp argue that decisions and actions made in response to persuasion are non-controlled and autonomous, assuming all other conditions of autonomy are satisfied (Faden and Beauchamp 1986, 262).

The authors argue only that the three proposed conditions are necessary conditions for autonomous action. They do not claim that they are sufficient conditions, and they openly entertain the possibility that other conditions, such as non-control by psychiatric conditions, addiction, and so forth, may be required for actions to be truly autonomous (Faden and Beauchamp 1986, 268). This is a different perspective on non-control as formulated above, which only requires independence from controlling influence by others. It is also important to note that Faden and Beauchamp's theory of autonomy does not require the conditions of understanding and non-control to be perfectly satisfied in order for an action to count as autonomous. Instead, they argue, the goal is for decisions and actions to be *substantially autonomous*. That is, they are intentional actions performed with substantial understanding and are substantially uncontrolled. Faden and Beauchamp do not set forth a specific threshold for substantiality, arguing that such precision is better afforded in particular situations rather than a general theory (Faden and Beauchamp 1986, 241).

What is Informed Consent?

Faden and Beauchamp distinguish between two senses of informed consent: informed consent as an autonomous authorization and informed consent as institutionally effective consent. Institutionally effective informed consent refers to informed consent obtained through procedures that adhere to the policies and rules defining a specific institutional practice. In this sense, informed consent is a legally or institutionally effective authorization, but it is not necessarily an *autonomous* authorization. As long as the person seeking consent satisfies the applicable institutional rules governing informed consent, the consent obtained from the patient is institutionally effective, regardless of whether the patient autonomously authorized the proposed intervention. By the same token, a patient may autonomously authorize a proposed intervention, but if the applicable rules for obtaining informed consent were not followed, the consent would not be considered institutionally effective or valid (Faden and Beauchamp 1986, 280-281).

There is no necessary connection between institutional rules that stipulate the features of informed consent and the logically necessary conditions derivable from the conditions for autonomous actions. Absent such a connection, the institutional rules are simply normative requirements that govern the practice of obtaining informed consent. An institutionally-required level of informational disclosure, for example, is not a condition that necessarily follows from the conditions required for autonomous action. A physician who is very familiar with an intervention could likely autonomously authorize another physician to perform that intervention on him without having the pertinent information, as defined by the institution, disclosed to him by that physician or another medical professional (Faden and Beauchamp 1986, 276). This is an important distinction because it highlights the ways in which the rules governing the institutional practice of informed consent can be erroneously conflated with the moral requirements essential to the respect for patient autonomy. Of course, this is not to say that such normative requirements are not practically useful, or even necessary, in given contexts. The point is that a violation of institutional rules is not necessarily a violation of an individual's right to autonomous action.

Faden and Beauchamp's formulation of informed consent in the first sense, as an autonomous authorization, attempts to transform the essential conditions of autonomous action more broadly into the logically necessary conditions for informed consent. They argue that just as consents and refusals are species of the broader category of actions, informed consents and informed refusals are species of the broader category of *autonomous actions*. Informed consent is, thus, not synonymous with autonomous action. It is rather a specific type of autonomous action, an autonomous authorization by patients or research subjects, whose conditions are not identical to those of the larger category of autonomous action (Faden and Beauchamp 1986, 277).

The idea of an informed consent suggests that a patient or subject does more than merely agree with, assent to, or comply with a proposal. In giving informed consent, the patient actively authorizes the proposed treatment. A patient may assent to a treatment without *authorizing* it. The assent may simply reflect the patient's submission to the physician's order. If this were the case, the patient would not be relying on his or her own authority in order to give the physician permission to pursue the proposed course of treatment. Hence, he is not authorizing the treatment (Faden and Beauchamp 1986, 278). Faden and Beauchamp offer the following definition of informed consent:

> An informed consent is an autonomous action by a subject or a patient that authorizes a professional either to involve the subject in research or to initiate a medical plan for the patient (or both). Following from the analysis of *substantial autonomy* . . . informed consent [in the sense of an autonomous authorization] is given if a patient or subject with (1) substantial understanding and (2) in substantial absence of control by others (3) intentionally (4) authorizes a professional (to do I) (Faden and Beauchamp 1986, 278).

Conditions 1-3 follow directly from Faden and Beauchamp's analysis of substantially autonomous actions more generally, but the fourth condition does not. This condition is what distinguishes informed consent as one kind of autonomous action (Faden and Beauchamp 1986, 278).

John Kleinig offers an interpretation that, while not explicitly referring to consent as an autonomous action, nonetheless helps to distinguish informed consent as simply one kind among the larger category of autonomous actions. He describes consent as a transaction in which consent to something is always given to another party or agent. It is a communicative act that alters the moral relations between two parties. When person A consents (to person B) to act X, person B is granted a moral right or entitlement to perform act X with respect to person A (Kleinig 2010, 5). In Kleinig's words, "B seeks from A either A's permission to do something or A's agreement to do something—something to which B had no moral right or entitlement" (Kleinig 2010, 6). Because consent is a communicative act, it requires that A signify it to B, though the conventions by which we recognize consent vary considerably across different contexts. Nonetheless, Kleinig argues, consent "will commonly be constituted by some gesture, word, or other recordable behavior that conventionally and contextually expresses it" (Kleinig 2010, 11).

There are two important elements of Kleinig's description of consent that help to clarify what is implied in Faden and Beauchamp's analysis. First, consent is always part of a transaction that is at least dyadic. To say that an agent, A, consented to an act, X, would be to give an incomplete account of an act of consent. The account must include at least another agent, B, to whom A gave consent to act X.[4] (Kleinig 2010, 5-6). This element is also present in Faden and Beauchamp's account, where, they contend, the patient must authorize a professional (i.e., another agent) to perform the proposed medical intervention. It may seem obvious that consent to a medical procedure involves more than one party, but it is worth highlighting because there is nothing inherent in any medical procedure that entails autonomously undergoing it must involve consent. A physician may, for example, autonomously (i.e., intentionally, with substantial understanding, and in substantial lack of control by others) perform a medical procedure on himself, but it would be mistaken to suggest that he provided consent to himself to perform the procedure.

The second important element of Kleinig's account is the notion that prior to obtaining A's consent to do or perform an act, B had no moral right or entitlement to perform that act. Because B has no right or authority to act until he is granted that right or

---

[4] According to Kleinig, A and B need not be individual persons. They may also be collective persons, whether as members of a particular class (e.g., the shareholders of a company) or as a collective unity (e.g., an orchestra). When the majority of shareholders communicates its agreement with a takeover offer, it consents to the takeover. Similarly, when an orchestra agrees to play in a certain location, it consents as a collective unit, regardless of whether the consent reflects the preference of each member (Kleinig 2010, 5-6).

authority by A (from whom he seeks consent) the act of consent alters the moral relations between A and B, granting him or her the right to proceed. That the actor initially lacks the right or authority to proceed with his proposed action is implied in Faden and Beauchamp's characterization of an authorization. In consenting to and, thus, authorizing a treatment, they argue, the patient assumes responsibility for what he or she has authorized and *transfers* the authority to perform the treatment to the physician (Faden and Beauchamp 1986, 280). The most important element of the authorization "is that the person who authorizes uses whatever right, power, or control he or she possesses in the situation to endow another with the right to act" (Faden and Beauchamp 1986, 280).

Extracting the essential conditions of informed consent offered by Faden, Beauchamp, and Kleinig, one could say that informed consent occurs when agent A autonomously transfers the moral right or authority to another agent, B, to perform a proposed act, X, which prior to the transfer, agent B had no moral right or authority to perform. Consent to a medical treatment, then, would occur when a patient autonomously transfers the moral right or authority to a medical professional to perform a proposed intervention, an intervention which the medical professional did not have the moral right or authority to perform prior to the transfer. Intrinsic to the consent transaction is a loyalty to the component parts of autonomous action: *agency* (i.e., intentionality and understanding) and *liberty* (i.e., non-control).

The liberty component entails a negative obligation on the part of the medical professional(s) to refrain from exerting controlling influences on the patient. The agency component represents a positive obligation on the part of the medical professional(s) to ensure patient understanding and intentionality. This is generally fulfilled by way of an adequate informed consent process in which the medical professional, among other things, discloses relevant information to the patient. The agency component also, however, represents certain capabilities of the patient. Regardless of whether the patient received adequate information and was situated in an environment free of controlling influences, he would not be able to act autonomously unless he had the capacity to understand the information to which he enjoys access. This concept is the inner psychological component of autonomy referred to as competence or decision-making capacity.

Autonomous Refusal of Treatment

The context in which a patient expressly refuses treatment is the same as that in which a patient consents to treatment. It consists of a dyadic transaction between the patient and the physician, and absent consent, the physician has no moral right or authority to perform the proposed treatment on the patient. Furthermore, an autonomous refusal, like informed consent, is a specific type of autonomous action, and its conditions are derivable from the necessary conditions for autonomous action more generally. Faden and Beauchamp’s model of autonomy would require an autonomous refusal to satisfy the three conditions necessary for autonomous action. An informed refusal would be a refusal given intentionally, with substantial understanding, and in substantial absence of control by others. If an informed refusal is analogous to informed consent, and perfectly equivalent to it in terms of its necessary conditions, then the individual, in refusing, would also be authorizing (or transferring the moral authority to) the other agent to do something.

Since a refusal concerns a proposed act in which the individual does not wish to participate, one might be inclined to argue that an authorization or transfer of moral authority would be made regarding the non-performance of the proposed act. If so, an informed refusal would occur when agent A autonomously transfers the moral right or authority to another agent, B, *not to* perform a proposed act, X, which prior to the transfer, agent B had no moral right or authority *not to* perform. This formulation is akin to a relief from a specific obligation. If agent B has no moral right or authority not to perform act X, then absent agent A's authorization of agent B's not performing act X, agent B has a moral obligation to perform act X. According to this formulation, informed refusal of medical treatment occurs when a patient autonomously transfers the moral right or authority to a medical professional not to perform a proposed intervention, an intervention which the medical professional had an obligation to perform prior to the transfer of the moral right or authority. Under this construal, it makes sense to describe an informed refusal as a *consent to no treatment* because the refusal is simply an autonomous authorization by which the physician obtains the moral authority that he would have otherwise lacked.

Because the initial circumstances of the dyadic transaction are *not* that the physician lacks the moral right or authority *not to perform* the proposed intervention but rather that he lacks the authority *to perform* it, the above formulation of informed refusal is incorrect. Given the initial lack of authority to treat, it is nonsensical to declare that an informed refusal authorizes the physician not to perform the treatment. He or she does not need the authority not to do something that he or she does not have the authority to do anyway. To say otherwise would be equivalent to declaring that a libidinous man needs the authority to refrain from having sex with an unwilling woman. Absent consent, both the physician and the libidinous man simply do not have the authority to pursue their proposed courses of action.

As noted above, a more appropriate formulation of an informed refusal would be derived from the initial conditions that constitute an informed consent transaction, specifically that the physician initially lacks the authority to perform the treatment. Stemming from these conditions, an informed refusal occurs when agent A autonomously *refuses* to transfer the moral right or authority to another agent, B, to perform act, X, which both prior to and after the refusal, agent B had no moral right or authority to perform. Under this formulation, the initial conditions of the encounter are not altered by agent A's refusal to authorize act X. Agent B does not gain or lose any moral right or authority with respect to act X as a result of agent A's refusal. Agent B did not have the authority to perform act X before the refusal, and he or she does not have the moral right or authority to perform act X after the refusal. Agent A's refusal is, thus, a preservation of the status quo with regard to agent B's right to perform act X. In fact, even if agent A does not expressly refuse to transfer the moral right or authority to agent B, the status quo is preserved, at least theoretically. This is not to say that the notions of implied, presumed, or surrogate consent are not morally relevant in certain situations. It is simply to point out that the theoretical foundations of informed consent and, thus, informed refusal lie in agent B's initial lack of moral right or authority to proceed with a proposed course of action.

Though an informed refusal requires that an individual refuse a proposed action *autonomously* (i.e., intentionally, with substantial understanding, and in substantial lack of control by others), and an autonomous refusal is sufficient to preserve the status quo with regard to agent B's initial lack of authority, this does not imply that an autonomous refusal is necessary to preserve the status quo. In fact, the above formulation of an informed refusal would be no different in its effect on the moral relations between agent A and agent B if it were devised as a refusal that was not necessarily informed, or even one that was uninformed. Consider the effect of the following reformulation on agent B's moral right or authority: a refusal occurs when agent A *autonomously or non-autonomously* refuses to transfer the moral right or authority to another agent, B, to perform act X, which both prior to and after the refusal, agent B had no moral right or authority to perform. This reformulation is faithful to the initial conditions of the dyadic transaction, and it does not alter the meaning of the term *autonomous* in autonomous action. Still, a non-autonomous refusal to authorize agent B to perform act X would be no different in consequence than an autonomous refusal.

If, as argued above, a refusal is not an authorization or transfer of moral authority to a physician, then the role the agency conditions (intentionality and understanding) play in providing moral validity to a treatment refusal seems somewhat unclear. An individual can refuse to consent without providing an informed refusal, and this does not affect the moral relations between the two parties. For example, suppose that an informed refusal requirement were imposed on decisions about sex. This would mean that an individual's refusal of a sexual encounter is only morally valid if the individual (1) with substantial understanding and (2) in substantial absence of control by others (3) intentionally refuses the sexual encounter proposed by the other party. The failure to satisfy any of these conditions would render a refusal non-autonomous and would, thus, mean that the individual's expressed refusal is not sufficient to prohibit the proposer of the sexual encounter from proceeding. This seems to be an indefensible position. If individuals are required to *intentionally* refuse sexual relations, or sufficiently *understand* the nature and consequences of their refusal, then this would change the normal understanding of consent to sexual relations, in which, regardless of other factors, consent must be signified in some way for the proposer of a sexual encounter to proceed.

The above formulation of a refusal suggests that the moral force is not necessarily in the refusal itself but the lack of the transfer of moral right or authority. If the effect of the refusal is nothing more than a preservation of the status quo of the moral relations between agent A and agent B, then a broader, more neutral representation can be used to signify agent B's lack of moral right or authority. Consider the following reformulation: the status quo of the initial moral relations between agent A and agent B is maintained if agent A *does not* transfer the moral right or authority to another agent, B, to perform act, X, which absent such a transfer, agent B has no moral right or authority to perform. This would include but not be limited to active refusals to transfer the moral right or authority. One need not *refuse* to authorize something in order to not authorize it.

One challenge to this view is that it incompletely characterizes the context of the *clinical* encounter, which can differ from non-clinical consent or refusal transactions. While the initial moral relations between agent A and agent B are contextually relevant, they are not the only contextually-relevant factors when a physician proposes a treatment to a patient. Another relevant factor in the physician-patient encounter is the physician's obligations derived from the principle of beneficence. Since promoting patient welfare is also an obligation bestowed upon a physician by her role (Beauchamp and Childress

2009, 205), it should be given due consideration alongside the physician's initial lack of moral authority to treat. When a physician enters the practice of medicine, she accepts the obligation to observe the standards of the profession, and failing to do so is considered a breach of her professional duty (Beauchamp and Childress 2009, 154). Although this obligation is significant, it still does not provide the physician with the moral authority to treat a patient. Rather, it (generally) only obliges her to propose the most appropriate treatment (or a range of treatments) according to her professional judgment and, if the patient authorizes the treatment(s), to carry it out competently.

If the patient does not authorize the treatment(s), despite the physician's proposal, then the physician has an obligation to refrain from proceeding with the treatment she recommended. Thus, the physician has a general obligation to promote the patient's welfare, but she only has the specific obligation to treat the patient if she has been granted the moral right or authority to do so. In order for her to fulfill her obligation to promote the patient's welfare, however, she must at least propose the treatment(s) which she deems most suitable for the patient's condition. In some situations, she might be required to do more than this, such as by seeking the moral authority to treat from a source other than the patient.

The above reformulation characterizes a large family of scenarios, and refusal (autonomous and non-autonomous) is just one of many scenarios within this family. Some of the other scenarios this *non-authorization* reformulation encompasses are: a non-autonomous individual who agrees to a proposed treatment but is unable to autonomously authorize it; an autonomous individual who agrees to a proposed treatment but, for reasons unrelated to his or her inner psychological capacities (e.g., inadequate information disclosure), is unable to autonomously authorize it; and an unconscious individual who is unable to agree or disagree with a treatment, or refuse or autonomously authorize it. If the reformulation is the end of the story regarding a medical professional's moral right or authority to proceed with a proposed intervention, then this could lead to some morally unpalatable results for the first and third scenarios. There is nothing innately challenging with the second scenario because it can be remedied by conducting a more thorough informed consent process in which adequate information is disclosed. After this, the individual would be able to autonomously authorize the proposed treatment. For the other two scenarios, absent instilling the aptitude for autonomous action, it would be impossible for the individuals to authorize the interventions. This would mean that the patients would have to forgo treatment because there would be no way for the physician to obtain the moral right or authority to treat. Of course, a patient's lack of autonomy does not prevent him or her from receiving needed treatment simply because the patient cannot autonomously transfer the moral right or authority to the physician to treat. In such situations, the physician has the moral obligation, derived from the principle of beneficence, to seek from another source the moral right or authority to treat the patient.[5]

Decision-Making Capacity and Surrogate Decision Making

In contemporary medical ethics, the requirement that a medical professional gain the moral right or authority to treat from the patient only extends to situations in which the patient has the capacity to autonomously transfer that moral right or authority. This *capacity* functions as a "gatekeeping concept" (Faden and Beauchamp 1986, 288), to use

---

[5] As I will argue later on, however, scenarios in which a non-autonomous patient actively refuses treatment are different from those in which the patient is unconscious or agrees to treatment but lacks decision-making capacity.

Faden and Beauchamp's term, meaning that for individuals without capacity, their authorization of an action is not needed in order for another person to perform that action. Faden and Beauchamp describe this as follows:

> If a person is *autonomous* and situated in a context in which consent is appropriate, it is a prima facie moral principle (derived from the basic principle of respect for autonomy) that informed consent should be sought from that person. By contrast, if a person is *nonautonomous* and situated in a context in which consent is required, it is a prima facie moral principle (*not* derived from the principle of respect for autonomy, but rather from beneficence) that some mechanism for the authorization of procedures or decisions other than obtaining the person's consent should be instituted (Faden and Beauchamp 1986, 288).

Although there are numerous instruments used to measure it (Grisso, Appelbaum and Mulvey, et al. 1995), clinical decision-making capacity is fundamentally the capacity for autonomous action. It involves, primarily, the capacity to understand, in Faden and Beauchamp's sense of the term, though, as some have pointed out, the capacity to communicate one's decision is also required (Buchanan and Brock 1989, 23).

The lack of capacity to understand pertinent information and the implications of one's actions indicates non-autonomy and, therefore, the lack of capacity to give informed consent. As such, capacity to understand is a precondition for actual understanding, where actual understanding is necessary for an autonomous authorization of a treatment (Beauchamp and Childress 2009, 120). According to some standards, lack of *appreciation* for the nature and consequences of a treatment or its alternatives can also be said to indicate incapacity. A patient may have a superficial understanding of the risks and benefits of a procedure but not appreciate how they apply to her situation (Berg, et al. 2001, 102). For instance, a patient may understand and believe that a particular medical regimen carries with it the risk of death, but she may, for any number of reasons, deny that she, herself, is at such risk if she agrees to the treatment. An even more stringent standard of capacity would examine the decision-making process itself. A patient may understand and appreciate the nature and implications of a treatment, but still not process the information presented to him in a logical manner. An authorization based on flawed reasoning would not be considered autonomous and would, therefore, not constitute morally, or even legally, valid informed consent (Berg, et al. 2001). Though *appreciation* and *reasoning* are presented as more stringent standards, there are really only surface-level differences between these standards and the concept of *understanding* offered by Faden and Beauchamp. If someone is unable to appreciate how a situation applies to her, or is unable to reason logically about it, then it is unlikely that she truly *understands* in the sense required for autonomous action.

Regardless of which standard of surrogate decision making (reviewed in the chapter one) is morally superior, in a particular situation or more generally, the principles underlying surrogate decision-making are the same as those which form the basis for informed consent. If a physician would like to perform some treatment on a patient and the patient agrees (or assents) but lacks the capacity to grant informed consent, then an appropriate surrogate would be able to consent on the patient’s behalf. The surrogate’s consent would be a stand-in for the patient’s deficient autonomy, and it would grant the physician the moral right or authority to proceed with his proposed treatment.

Suppose, however, that a physician would like to perform some treatment, but the patient expresses his refusal to allow the physician to proceed. Suppose, further, that the patient clearly lacks the capacity to understand, appreciate, and/or reason logically about the nature and consequences of foregoing the proposed course of treatment. The expectation that refusals of treatment be “informed” inescapably leads to the conclusion that the patient’s refusal is not autonomous and his wishes (not to undergo treatment) need not be respected. The physician, however, would still not have the moral right or authority to proceed with treatment. Therefore, the consent of an appropriate surrogate decision-maker would be sought. Presumably, an *informed* refusal by the surrogate would be respected without dispute. The competent surrogate would be refusing to grant the physician the moral right or authority to treat the incapacitated patient. In this case, there is no tension because the refusal of the surrogate coincides with the wishes of the patient. On the other hand, if the surrogate decides to grant consent, and the consent conflicts with the expressed wishes of the patient (i.e., the patient expresses refusal), then the question is whether the physician has the moral right or authority to proceed. To answer this question, one must determine whether the surrogate has the right to grant the physician such a moral right or authority in spite of the patient’s refusal. To do this, it is first necessary to determine whether an individual’s active refusal contains an inherent moral significance that is independent of his or her decision-making capacity (i.e., capacity to act autonomously) and is, thus, a distinct form of *non-authorization*. Below, I will attempt to show that an active *refusal* does, indeed, contain an inherent moral significance that is not contingent upon capacity, and that this moral significance distinguishes it from the other types of non-authorization. If this attempt is successful, then a qualification of the *autonomy as a gatekeeper* concept proposed by Faden and Beauchamp is in order.

Does Consent of a Surrogate Nullify a Patient's Active Refusal?

There are numerous situations outside of the context of medical treatment in which individuals who lack the decision-making capacity nonetheless retain the right to refuse. Below, I will discuss a few examples. Before moving forward, it is important to note the difference between in*competence* and in*capacity*. Incapacity denotes the inability to make certain decisions autonomously. Incompetence is a legal construct designating one as not having the capacity and, therefore, the legal right to make those decisions (Berg, et al. 2001, 95-96). Thus, the commonly cited distinction in medical ethics is that competence relates to a legal judgment whereas capacity relates to a clinical one (Appelbaum, Assessment of Patient's Competence to Consent to Treatment 2007). This distinction is clear enough, but it is slightly inaccurate. In fact, a person either does or does not have capacity and its presence is not contingent upon the judgment of a clinician. The clinical judgment is simply a way of determining whether the person has capacity.

The first example concerns refusals to participate in research. In research contexts, as in clinical contexts, when a cognitively-impaired adult lacks the capacity to provide informed consent to participate in research, informed consent must be obtained by a surrogate decision maker (a legally authorized representative). Although the federal regulations governing research do not specifically require that adults without capacity assent to participation, various organizations and institutional policies respect an individual's refusal (or dissent), whether or not the surrogate consents on the individual's behalf. The Alzheimer's Association, in a consensus recommendation to institutional review boards and investigators, endorses the following approach:

> If the participant is capable of providing affirmative agreement to participate, the participant should be informed in the presence of the proxy that he or she is about to be enrolled in a research study. The procedures, risks, benefits, and alternatives involved should be explained in a simple fashion. The participant should then be asked if he or she agrees to be in the research, and the response should be recorded.
>
> If the participant is incapable of providing affirmative agreement to participate, then assent (or dissent) should be judged behaviorally based on cooperativeness with study procedures (e.g., does he or she refuse to have blood drawn, take pills, lie still for an imaging study?). Dissent for any study-related procedures should be respected, and consistent dissent may be a basis for removal from the research study (Alzheimer's Association 2004).

In a survey of Alzheimer Disease Centers funded by the National Institute on Aging, of the institutions that had created their own policies regarding research with cognitively impaired adults, four out of five policies stated that a subject’s dissent or unwillingness to participate must be honored (though two of these allow for dissent to be overridden by judicial action) (Cahil and Wichman 2000). The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, concerning research that holds out the prospect of direct benefit to institutionalized, “mentally infirm” individuals, recommends that an individual’s refusal of such research should not be overridden unless a court specifically authorizes participation and the intervention is only available within the context of the research (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1978, 8-14). However, “in jurisdictions that grant institutionalized individuals an unqualified right to refuse therapy, their objection to participation in research will be binding” (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1978, 14). For research that is more-than-minimal risk, but does not hold out the prospect of direct benefit, the National Commission holds that a potential subject’s refusal should be binding, regardless

(National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1978, 19)

Although the National Commission's recommendations allow for potential subjects' refusals to be overridden by the courts when there is a prospect of direct benefit that is only available through participation, the fact that refusals to participate in studies that do not fit into this category (i.e., those with no prospect of direct benefit) are respected without exception suggests that refusals are morally significant irrespective of decision-making capacity (at least in these cases). This moral significance distinguishes an active refusal from other forms of *non-authorization* in which the subject is not capable of assenting to or refusing participation. In these instances, the National Commission contends, the "absence of objection should be sufficient to permit participation in research that is relevant to the subject's condition and presents no more than minimal risk" (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1978, 10). Regarding research protocols that are potentially beneficial, I would argue, first, that the National Commission's position on overriding subjects' active refusals simply reflects the sentiment regarding beneficial *treatment* in mainstream medical ethics, which is what I am attempting to appraise in this analysis. Secondly, the requirement that the refusal be overridden by a court authorization does not suggest that the subject's incapacity disqualifies him or her from making a decision, as the *autonomy as a gatekeeper concept* maintains. Rather, it implies that the subject's objection to participation is noteworthy, but it can be overridden by a court authorization. That a court authorization is required means that an individual's lack of decision-making capacity is not an automatic justification for enrolling her against her wishes.

The next examples refer to minors, where it could be argued that although they are not legally *competent* to consent to certain activities, they still may have the *capacity* to do so. Such an observation is correct. Competence and capacity do not correlate perfectly in minors. As such, some commentators claim that age is an arbitrary determinant of competence and minors' actual abilities should be taken into account (Alderson 2007). Nonetheless, it is reasonable to assume that some, perhaps most, minors are both legally incompetent and decisionally incapacitated due to their age. Since this exploration is concerned with capacity, the examples below should be taken to apply only to these cases. Also, it should be noted that although the examples refer to minors, the intent behind the analysis is to show that *nonautonomous* refusals of any kind, not just those expressed by minors, are morally significant.

Consider the case of minors and tattoos. Some states prohibit the tattooing of minors in all cases except those under the direct supervision of a licensed physician or osteopath (presumably for medical purposes). In such states, *no one* may consent on behalf of a minor to a tattoo for cosmetic purposes (O.C.G.A 16-5-71 (2012) enacted by Ga. L. 1987). In other states, written parental consent is required before a person may tattoo (or brand or pierce) a minor's body (Section-17A-2 n.d.). Such statutes, asserting minors' legal incompetence to consent, imply a lack of capacity in some respect. The minor lacks the capacity to grant the tattoo artist the moral right or authority to perform a tattoo on his body. The laws implicitly recognize that the child does not possess the requisite autonomy to understand, appreciate, and reason about the nature and consequences of receiving a tattoo. He is, therefore, unable to provide legally and, more importantly for our purposes, morally valid consent. The consent of the parent, however, at least according to the states which allow tattooing of minors, would be morally valid. It would be sufficient to grant the tattoo artist the moral right or authority to perform the tattoo. It would be sufficient in the same way as a parent's consent to treatment on behalf of a child would give the physician the moral right or authority to proceed. Still, the parent's authorization does not neutralize a minor's right to refuse a tattoo, quite the opposite. Despite a minor's lack of capacity to consent to a tattoo, and a parent's capacity to consent on his behalf, the tattoo artist does not have the moral authority to proceed without the assent of the minor. Of course, it could be countered that although the minor does not have the capacity to understand, appreciate, and/or reason logically about the nature and consequences of getting a tattoo, he certainly has the capacity to refuse one. This may, in fact, be true, but what if the child was extremely intellectually disabled? What if he was also schizophrenic and thought that the tattoo artist was a government agent planning to inject mind-control ink into his body? In such cases instances, the minor clearly would lack the capacity to refuse (in terms of traditional *decision-making capacity*), but the parent still would have no right to grant the tattoo artist the right or authority to tattoo the child. This example may seem unimpressive because of its implausibility. It is, indeed, highly unlikely that such an egregious moral violation would occur. Nevertheless, it is the egregious nature of the act, and the fact that the tattoo serves no apparent purpose, that illuminates the moral significance of the child's indisputably non-autonomous refusal.

Next, consider state laws mandating parental involvement in minors' abortion decision. As of January 1, 2015, 38 states require parental involvement in a minor's decision to have an abortion: 21 require parental consent only (3 require consent of both parents), 12 require parental notification only, and 5 require both parental consent and notification. Thirty-six of the states requiring parental involvement have a judicial bypass procedure in which a child may obtain approval from a court to have an abortion without parental involvement (Guttmacher Institute 2015). The Supreme Court, in *Bellotti v. Baird* in 1979, held that under such legislation, parents could not veto a minor's decision, but their involvement could be required so as to ensure better decision-making. A minor whose parent(s) refused to provide consent could, therefore, seek permission from a judge through a judicial bypass hearing. In a bypass hearing, a judge may authorize a minor's abortion after determining either that the minor is sufficiently mature to make her own decision *or* that the abortion is in her best interests (Maniam 2012). This is somewhat complicated for the purposes of this analysis because underlying the law is an amalgam of concerns: legal competence, decision-making capacity, parental consent, and non-parental (judicial) surrogate consent. In a judicial bypass hearing, the former determination, that the minor is mature enough to decide on her own, is essentially a judicial declaration of the minor's legal competence based on evidence of her decision-making capacity. It is affirming her observable autonomy by granting a corresponding legal status of competence. The latter determination is more similar to surrogate consent in the medical context. Presumably, the determination is made only for minors without sufficient decision-making capacity; otherwise, the former determination could be made. Given the minor's lack of capacity, the judge's authorization is what grants the physician the moral right or authority to perform the abortion. Of course, it could be argued that although the judge has the legal authority to authorize the physician to perform an abortion, his authorization would not be morally transformative because he does not have the moral authority to do so in the face of parental refusal. Assuming he *does* have the moral authority to authorize the abortion, he still would not have the moral authority to authorize an abortion against the wishes of the child just because she lacked the capacity to understand, appreciate, and reason logically about the consequences of carrying a pregnancy to term. This is implied by the parental involvement laws themselves, which provide for judicial hearings only for minors *seeking* abortions. The underlying rationale is, therefore, to protect minors who want abortions but lack the capacity to autonomously authorize them. Otherwise, it would seem defensible to hold hearings for *all* pregnant minors whose decisions conflict with those of their parents. The judicial determination would involve a judgment of whether the minor has the capacity to decide on her own. If the judge thinks not, he would determine if an abortion was in the minor's best interests. If he decided it was, he could authorize the abortion procedure whether the minor wanted it or not. This would be a morally egregious violation of the minor's bodily integrity, but it would be permissible so long as the *autonomy as a gatekeeper concept* applied to both consent and refusal.

Capacity and the Consequences of Refusal

One answer to these examples could be that there is an essential difference between decisions about necessary medical treatment and decisions about non-therapeutic research, tattoos, or abortion. Refusing needed medical care, an argument would go, could carry negative consequences whereas this is not likely to result from the other, less serious cases of refusal. Factual accuracy of such a claim aside (consider the consequences of teenage pregnancy), the principal assertion is that one must have a greater capacity to understand, appreciate, and reason about the *nature* and *consequences* of refusing necessary medical treatment than he would need to understand, appreciate, and reason about "less serious" decisions. This position views the capacity to refuse to be acted upon (or to consent to be acted upon) on a sliding-scale: the more serious the consequences of refusal, the more rigorous the standard of capacity required (Appelbaum 2007). A child or mentally ill or disabled adult may, for instance, have the capacity to refuse an assessment of his vital signs at a routine doctor's visit, but he may not have the capacity to refuse a life-saving surgery. As Beauchamp and Childress note, such a strategy is only useful if the level of evidence required for determining capacity, not the level of capacity itself, rises with the seriousness of the decision. We, therefore, need to distinguish between a patient's capacity and how we determine that capacity (Beauchamp and Childress 2009, 116-17), for "no basis exists for believing that risky decisions require more ability at decision making than less risky decisions" (Beauchamp and Childress 2009, 117). Their approach recognizes that the complexity of the decision rather than its likely consequences is a more appropriate determinant of the required level of capacity (Beauchamp and Childress 2009, 117).

Still, the approach does not address cases in which the nature of the decision demands only a low level of evidence of capacity, but there is unavoidable and irrefutable evidence of incapacity. Consider, again, the schizophrenic minor who refuses to be tattooed. Because the risks of refusing a tattoo are arguably non-existent, the decision to refuse one would seemingly require a less-than-stringent capacity assessment.

Nonetheless, the teen’s expressed belief that the tattoo artist is planning to inject him with mind-control ink demonstrates a complete lack of understanding, appreciation, and ability to reason logically. He lacks capacity according to *any* standard. Regardless of his lack of capacity, however, it is not permissible to tattoo him against his expressed wishes. His refusal, thus, carries moral weight independent of his so-called *capacity* to autonomously refuse. Similarly, there are no evident risks associated with refusing to participate in a research study that offers no direct medical benefit. Regardless, an individual who does not wish to participate because of irrational fears or a manifest inability to appreciate or reason about the implications of participation lacks the capacity to refuse participation. Even so, his non-autonomous refusal is respected.

The case of minors and abortion is even more problematic for a claim that the consequences of a refusal should determine the required evidential level of capacity. If it is impermissible to force a minor with limited decision-making capacity to have an abortion, then the consequences of carrying the pregnancy to term and the complexity of such a decision are irrelevant to the respect-worthiness of her refusal. Besides, deciding to forego an abortion and enter motherhood at an early age *is* a significant decision. Although the consequences of maintaining a pregnancy as a teenager are largely social (assuming no medical reasons make the pregnancy uniquely risky), it is still difficult to make the case that there is a relevant difference in terms of capacity between the nature and consequences of refusing an abortion and the nature and consequences of refusing more urgently needed medical care. The only discernible difference, as noted above, is that one carries largely social consequences while the other carries “medical” consequences. This distinction breaks down once one considers how these types of consequences are inextricably linked and, perhaps, categorically equivalent. For instance, one *medical* consequence of refusing to control one's diabetes could be the need to amputate a leg. This *medical* consequence has social implications once the person leaves the clinic and must live as a disabled person. Living without a leg is the more lasting consequence, as is raising a child as a teenager. Thus, there is no relevant difference between the nature and consequences of refusing needed medical care and, say, refusing an abortion.

Should We Respect All Refusals?

So far, we have seen that even when an individual lacks capacity, his or her refusal to transfer the moral right or authority to perform an act is still morally significant. An active *refusal* to transfer the moral right or authority to a physician can, thus, be distinguished from other types of non-authorization in which autonomy acts rightly as a gatekeeper to the right to make one's own decisions. As such, the consent of an appropriate surrogate, necessary for the physician to have the moral right or authority to proceed with treatment, does not necessarily nullify that moral significance. This perspective eschews the view of the patient's incapacity as an autonomy vacuum to be filled by a surrogate, whose decision is as authoritative as the patient's would have been had he enjoyed the capacity to make it. It acknowledges that there is a genuine moral tension between the expressed wishes of the patient and the decision of the surrogate, and that the tension cannot be relieved by opining that the patient *is not autonomous anyway*.

In the next chapter, I will explore this tension. I will argue that it is present because the source of the moral significance is the individual's negative liberty interests, which, although inescapably fundamental to the notion of autonomy, have a justification independent of it. Using the independent justification for negative liberty, I will show that the structure of autonomy described in chapter one needs modification. By doing so, it should become clear that the moral authority of a surrogate to consent to an unwilling patient's treatment must be justified on grounds other than the patient's lack of decision-making capacity.

CHAPTER 3

As I argued in the first chapter of this thesis, autonomy in medical decision making assumes a particular structure which justifies compulsory treatment of patients who lack medical decision-making capacity. If a patient lacks *agency* in the form of medical decision-making capacity, then his right to *liberty*, or his right to act without the controlling influences of others, is also forfeited, and his expressed wishes at the time of the medical interaction are supplanted by the decision of a surrogate decision maker. This structure imposes a logical relationship between the concepts of agency and liberty in which agency is a definitional necessity of liberty. Since agency is a necessary condition for liberty, an individual does not have the right to act in the absence of the controlling influences of others if he or she does not possess the contemporary notion of agentive capacity. This relationship is the crux of the contemporary justification for compelling patients who do not possess medical decision-making capacity to undergo treatment against their wishes.

Of course, the logical relationship between agency and liberty is not wholly problematic. To be able to act upon one’s wishes without the controlling influences of others, one must, at the very least, be able to have wishes and be able to act upon them. If an individual is not capable of having wishes and/or acting upon them, then discussions regarding his right to act without external interference by others are meaningless. Such a minimal conception of agency, however, has completely different implications for the concept of liberty than the notion of agency proffered by contemporary medical ethics. The relationship between the minimal conception of agency and liberty is one of logical necessity because one simply cannot be coerced against his wishes if he does not have wishes. Further, he simply cannot act according to his own wishes if he lacks the capability of acting.

The relationship between medical agency and liberty is different. To possess medical agency is to possess medical decision-making capacity, which requires capabilities beyond the mere possession of wishes and the ability to act upon them. One's wishes regarding medical treatment must derive from certain cognitive capacities that allow one to understand a treatment or procedure, deliberate about its risks and benefits, and decide whether to undergo the treatment or procedure based upon the rational deliberation (Beauchamp and Childress 2009, 114). The lack of such capacities, in the realm of contemporary medical ethics, disqualifies one from his or her entitlement to liberty, or right to refuse medically-warranted treatment.

The primary difference between a minimal version of agency, the mere ability to have wishes and act upon them, and agency in the medical sphere lies in the type of relationship each version has with the concept of liberty. The relationship between the minimal version and liberty concerns the *ability* of the individual to act rather than his or her right to act. If liberty requires, at the very least, the right to act on one's wishes without external control, then, as stated above, one must have the capability of doing so in order to exercise his or her liberty, though she may still have the right to liberty even though she is unable to exercise it. Agency and liberty are only connected in that an individual who cannot have wishes or act cannot have his wishes or actions infringed upon by others. The relationship between medical agency and liberty, though grounded in the abilities associated with decision-making capacity, concerns the *right* of the individual to act on his or her wishes rather than his or her ability to do so. If an individual does not possess medical decision-making capacity, he may still have certain wishes and the ability to act upon them, but, regardless, he does not have the right to act upon them because medical decision-making capacity is a necessary condition for liberty. The primary difference, then, is that medical agency is a normative requirement for the entitlement to liberty, whereas a minimal version of agency is simply a requirement inherent to liberty's exercise.

Isaiah Berlin and Two Concepts of Liberty

Thus far, the concepts of agency and liberty have been discussed as two conditions for medical autonomy. An autonomous medical decision is one in which the patient possesses medical agency in the form of medical decision-making capacity and therefore enjoys the liberty to make his own decision. Because he satisfies the criteria for medical agency proffered by contemporary medical ethics, he may not be coerced to undergo medically-warranted treatment should he decide against it. His right to medical liberty, then, is dependent upon his possession of agency. Put another way, coercing the patient to undergo treatment, that is, depriving him of his right to act according to his expressed wishes, is justified by his lack of decision-making capacity.

These two concepts, liberty and agency (as articulated thus far), mirror the concepts of *negative* and *positive liberty* described by Isaiah Berlin in his influential essay, *Two Concepts of Liberty*. Berlin's concept of *negative liberty* is similar to the concept of liberty offered by Beauchamp and Childress. Likewise, Berlin's concept of *positive liberty* is tantamount to the concept of agency (medical agency) as it has been used thus far. Berlin did not claim that these two notions were exhaustive of the philosophical concept of liberty. Rather, he claimed that they were two dominant conceptions of liberty and have occupied a large portion of the philosophical history behind it. (Berlin 2000, 193-194). The following sections will discuss Berlin's views on these two notions of liberty and how they apply to medical autonomy. Before proceeding, a preliminary note is in order. Berlin uses the terms liberty and freedom interchangeably, and he makes it clear that he is doing so (Berlin 2000, 194), so use of either of these terms should be read as referring to the same concept.

Negative Liberty

Negative liberty, according to Berlin, is "simply the area within which a man can act unobstructed by others" (Berlin 2000, 194). The central obstacle to negative liberty is coercion, and a measurement of an individual's negative liberty is the degree to which he or she is free from coercion. Berlin provides the following:

> If I am prevented by others from doing what I could otherwise do, I am to that degree unfree; and if this area is contracted by other men beyond a certain minimum requirement, I can be described as being coerced, or, it may be, enslaved (Berlin 2000, 194).

Berlin is clear, however, that for interference with an individual's activities to be coercion, the source of the interference must not come from non-human sources. He maintains that "you lack political liberty or freedom only if you are prevented from attaining a goal by human beings," and that "mere incapacity to obtain a goal is not lack of political freedom" (Berlin 2000, 194).

Since negative liberty requires only that the individual be free from the coercive actions of other human beings, it relies on a minimalist conception of agency as described above. It does not require the individual to possess certain internal capacities or values for him to be entitled to non-interference by others. Rather, negative liberty presumes an entitlement to non-interference from others and imposes an obligation to refrain from obstructing an individual’s activities. John Stuart Mill defends negative liberty in his essay *On Liberty*:

> The object of this Essay is to assert one very simple principle . . . That principle is, that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right (Mill 1989, 49).

Positive Liberty

Liberty in the positive sense is the freedom accompanied by being one’s own master. It represents freedom from “nature” or one’s “own ‘unbridled’ passions” (Berlin 2000, 204). It involves, among other things, the “higher,” rational self achieving mastery over the lower self, the self which is dominated by irrational desires and impulses (Berlin 2000, 204). This idea of an individual having two selves, a rational, ideal self and an empirical self, is fundamental to positive liberty. Regarding these two selves innate to this notion of liberty, Berlin provides the following:

> This dominant self is then variously identified with reason, with my ‘higher nature’, with the self which calculates and aims at what will satisfy it in the long run, with my ‘real’, or ‘ideal’, or ‘autonomous’ self, or with my self ‘at its best’; which is then contrasted with irrational impulse, uncontrolled desires, my ‘lower’ nature, the pursuit of immediate pleasures, my ‘empirical’ or ‘heteronomous’ self, swept by every gust of desire and passion, needing to be rigidly disciplined if it is ever to rise to the full height of its ‘real’ nature (Berlin 2000, 204).

As can be seen, fundamental to this version of liberty is a higher form of agency than that required by negative liberty, the minimalist version. Positive liberty requires certain essential capacities or conditions, which may vary according to the particular form of positive liberty being endorsed, that are, by definition, required for an individual to be considered free. The common assumption underlying this line of thought, according to Berlin, is “that the rational ends of our ‘true’ natures must coincide, or be made to coincide, however violently our poor, ignorant, desire-ridden, passionate, empirical selves may cry out against this process. Freedom is not freedom to do what is irrational, or stupid, or wrong” (Berlin 2000, 219).

The Relationship Between Positive and Negative Liberty

The potential danger of the notion of positive liberty, according to Berlin, is that it divides the individual into two selves: the true, or rational self, and the empirical self, which is subject to the irrational passions and desires that need to be controlled or contained. Once this metaphorical bifurcation of the self has occurred, he argues, the door is open to the infringement upon one’s empirical wishes and desires in the name of his ‘true’ self, or his own freedom (Berlin 2000, 204-205). He argues:

> What, at most, this entails is that [the individual] would not resist me if they were rational and as wise as I and understood their interests as I do. But I may go on to claim a good deal more than this. I may declare that they are actually aiming at what in their benighted state they consciously resist, because there exists within them an occult entity – their latent rational will, or their true purpose – and this entity, although it is belied by all that they overtly feel and do and say, is their ‘real’ self, of which the poor empirical self in space and time may know nothing or little and that this inner spirit is the only self that deserves to have its wishes taken into account. Once I take this view, I am in a position to ignore the actual wishes of men or societies, to bully, oppress, torture them in the name, and on behalf, of their real selves, in the secure knowledge that whatever is the true goal of man (happiness, performance of duty, wisdom, a just society, self-fulfillment) must be identical with his freedom – the free choice of his ‘true,’ albeit often submerged and inarticulate, self (Berlin 2000, 205).

George Crowder calls Berlin’s argument the “inversion thesis,” the idea that the notion of positive liberty allows the concept of liberty to be inverted into its very opposite. Coercion can be justified because the coerced actions or desires are purported to be more consistent with liberty than the individual’s actual wishes. Crowder points out that there is a strong undercurrent in Berlin’s thesis that the logic of positive liberty ought to make us suspicious because the idea itself exposes it to the potential for authoritarian corruption (Crowder 2004, 69). Although it would be somewhat dishonest to characterize the mainstream justification of coerced treatment as a form authoritarian corruption, the rationale for coerced treatment is nonetheless parallel to that of coercion in the name of positive liberty.

Medical Decision-Making Capacity as Positive Liberty

The contemporary moral justification for the compulsory treatment of patients who lack medical decision-making capacity rests on an idea of medical autonomy that fits within the realm of positive liberty. Just like Berlin’s notion of positive liberty, the theory of *medical agency* in the form of medical decision-making capacity relies on a bifurcation of the individual into two selves: the empirical self that presents to the clinic with whatever internal capacities the individual may possess and the “ideal” or “true” self, which is characterized by the capacities inherent to medical agency. When the patient’s empirical self is not in harmony with the true, ideal, autonomous self, his empirical wishes do not derive from the idealized form of decision-making endorsed by contemporary medical ethics. The patient, therefore, lacks medical agency and is not free in the positive sense.

Since the patient is not free in the positive sense, he has no right against coercion since the structure of autonomy in medical ethics requires one to possess medical agency (positive liberty) in order for him to enjoy medical liberty (negative liberty). That is why the compulsory treatment of an individual who lacks medical agency is not considered a violation of his autonomy: because he is not autonomous anyway. The *gatekeeper* role (Faden and Beauchamp 1986, 288) played by medical agency, or decision-making capacity, relieves the medical professional from any consideration of the patient's expressed wishes. Hence, the patient, by definition, does not have the right to have his wishes respected. Such an arrangement is what concerned Berlin. By enforcing a particular view of freedom, the basis of which is a set of internal capacities necessary to be considered free, one is able to justify coercion of an individual by citing his deficiency in the capacities that were prescribed him. In other words, one is able to ignore the expressed or "empirical" wishes of an individual patient and use the language of autonomy to justify coercing him.

Berlin is not wholeheartedly against coercion for the good of the individual (Berlin 2000, 204). Rather, he is against the imposition of certain philosophical ideals on individuals in the name of their own freedom. In the medical realm, those ideals would be the agency required in the form of medical decision-making capacity to refuse medically-warranted treatment. The non-autonomous patient, with his "poor earthly body and foolish mind" (Berlin 2000, 206) expressly rejects the needed treatment, but his empirical body is not truly free, so he is not really being coerced. Instead, his higher self has willed it, "not indeed consciously, not as he seems in everyday life, but in his role as a rational self which his empirical self may not know" (Berlin 2000, 205). What Berlin is criticizing is a particular justification for coercive action, the view that the tension between the expressed desires of an individual and a specific conception of his own good can be relieved by opining that the individual, in his current empirical configuration, is not free anyway. “Enough manipulation with the definition of man,” he points out, “and freedom can be made to mean whatever the manipulator wishes” (Berlin 2000, 206). It is, thus, not the coercion in itself that Berlin is distrustful of, but the imposed ideal of freedom that is used to justify it. As Christman put it, for Berlin, “to label as ‘freedom’ the mastery of the ‘lower’ desires by the higher capacities of morality and virtue, not to mention by the supposedly superior wisdom of a general will, marked a treacherous tilt toward the justification of centralized power under the guise of moral superiority” (Christman, Saving Positive Freedom 2005).

Objections to Berlin’s Distinction

Gerald MacCallum rejects Berlin’s position that there is a useful distinction between negative and positive freedom (or liberty). Rather, he argues, there is only one concept, and it is mistaken to characterize freedom as either one of two dyadic relations, “freedom from” (negative liberty) and “freedom to” (positive liberty). Instead, he maintains freedom is always a triadic relation in which an *agent* is free *from* some constraint *to* do (or not do) something (MacCallum 1967). All discussions of freedom can be fit into the format “x is (is not) free from y to do (not do, become, not become) z,” where “x ranges over agents, y ranges over such ‘preventing conditions’ as constraints, restrictions, interferences, and barriers, and z ranges over actions or conditions of character or circumstance” (MacCallum 1967, 314). On one hand, this view seems to collapse Berlin’s distinction between positive and negative freedom into one concept. If anything that hinders liberty can be conceptualized as a *constraint*, whether it is internal or external, then there is no useful distinction between negative liberty and positive liberty. It does not matter, according to MacCallum’s conceptualization, whether an individual's actions are interfered with by another agent or by his or her own inner psychological capacities. Both are instances in which the individual's liberty is constrained by something.

On the other hand, it is possible to view Berlin's distinction as consistent with MacCallum's formulation. Although MacCallum suggests Berlin is arguing that negative and positive liberty are exclusive of one another, Berlin's distinction does not exclude the possibility of common ground between the two concepts. Negative and positive liberty can be seen as different aspects of an overarching concept of liberty, but aspects which are, nonetheless, incommensurable. Instead of sharing a single essence, as MacCallum seems to imply, they could be seen as belonging to a family of concepts with an underlying commonality (Crowder 2004, 78). Even if all impediments to liberty can be viewed simply as constraints, this does not preclude a conceptual distinction between internal and external constraints. There are, at the very least, experiential differences between the two, a crucial point which will be taken up in a later section.

Christman's Content-Neutral Conception of Positive Liberty

John Christman argues that a positive concept of liberty need not be amenable to the so-called authoritarian corruption against which Berlin warned. Whereas Berlin emphasized the historical notions of positive liberty that opened the door for coercion because the coerced desires or actions supposedly conformed with reason more so than the individual's own, Christman argues that the concept of positive liberty does not have to rely on such strict conditions. It is not the content of the individual's desires, he argues, but the procedures by which his or her desires are formed that constitute positive liberty. According to Christman, a person P is positively free with regard to some desire D if:

> 1. P was in a position to reflect upon the processes involved in the development of D;
> 2. P did not resist the development of D when attending to this process of development, or P *would not have* resisted that development had P attended to the process;
> 3. The lack of resistance to the development of D did not take place (or would not have) under the influence of factors that inhibit self-reflection (unless exposure to such factors was autonomously chosen, in which case that choice had to be made without such factors); and
> 4. The judgments involved in this self-reflection, plus the desire set that results, are minimally rational for P (Christman, Liberalism and Individual Positive Freedom 1991, 347).

Christman elaborates on the fourth condition and its requirement of minimal rationality. Traditional accounts of positive liberty, he argues, are laden with declarations connecting “true” liberty with the demands of reason. The question, he maintains, is to what extent must the judgments involved in the self-reflection demanded by positive liberty be rational, or in what *sense* must they be rational? The criteria for rationality vary, and they can range from the requirement of consistency between beliefs and desires, to requiring the choice of the most effective means to achieve one’s ends, to having sufficient evidence for the beliefs upon which one’s desires depend. All accounts of rationality, however, can be put into one of two categories: “internalist” or “subjective” accounts and “externalist” or “objective accounts.” For an internalist account, the criterion by which an action is considered rational is dependent only on those beliefs and desires that are “internal” to the agent. The relation of those beliefs and desires to the external world (i.e., their accuracy) is not considered. It is usually demanded, Christman maintains, that the internal beliefs (upon which the conditional desires are based) are consistent and the desires are consistent. By contrast, the externalist account of rationality requires that an agent have adequate objective evidence to justify his beliefs, and that his desires be based on these beliefs. An even more stringent version of the externalist condition is one which requires the agent to conform his desires to the correct *values* as well as factual external evidence. Christman summarizes the distinction between internalist and externalist accounts of rationality as follows: “the internalist would only demand that a person acts for *reasons* (perhaps ones which meet some requirement of consistency), while the externalist demands that the free agent must act in accordance with *reason*, where that includes knowledge of the truth, both about the world as well as morality” (Christman, Liberalism and Individual Positive Freedom 1991, 350).

Christman defends the minimal, internalist account of rationality for the development of the desires of a positively free, autonomous person. This means that individuals whose actions are based on inconsistent beliefs or inconsistent desires are positively unfree. Christman does note that no one likely has completely consistent beliefs and desires, so the requirement is actually that there be a lack of *manifest* inconsistencies. He does not offer a point at which beliefs or desires should be considered manifestly inconsistent, but I suppose a line could be drawn, at least in theory. There is no requirement, however, that the beliefs in question fit the external (objective) facts, and there is, similarly, no requirement that the brute desires be appraised on the basis of their rationality (Christman, Liberalism and Individual Positive Freedom 1991).

Christman claims that this conception of positive liberty answers Berlin’s critique that proponents of positive liberty can justify interference with others’ actions by claiming the coercion is consistent with liberty. No one, he argues, will be in an epistemic position to justify interference on the basis of failed rationality of the internalist type. To do so, the interferer would have to know more than the agent about the internal structure of his desires and beliefs, and judge them to be inconsistent. This, according to

Christman, is practically impossible (Christman, Liberalism and Individual Positive Freedom 1991). Christman even argues that requiring an external *evidence* condition for one's beliefs would only allow for interference in a narrow range of cases. The cases would include, for instance, those in which the interferer has access to more factual information than the agent and where the information is indisputable and the agent had reasonable access to it. Interfering with an agent's actions under such circumstances is justifiable, Christman says, because "to act unwittingly is not to act freely. And if I interfere with your unwitting actions I do not disrupt your self-government in any meaningful way" (Christman, Liberalism and Individual Positive Freedom 1991, 355). Further, he continues, "most writers in the liberal tradition accept this as neither paradoxical nor pernicious" (Christman, Liberalism and Individual Positive Freedom 1991, 355).

If these less stringent conditions of positive freedom are accepted, and the notion that freedom requires adherence to the *correct values* is rejected, then what results is a content-neutral, autonomy-based conception of positive freedom. Christman defends this view as follows:

> There are good theoretical reasons for a content neutral conception. For any desire, no matter how evil, self-sacrificing, or slavish it might be, we can imagine cases where, given the conditions faced, an agent would have *good reason* to have such a desire. That is, there may be many cases where I freely pursue a strategy of action that involves constraining my choices and manipulating my values. But if this is part of an autonomous pursuit of a goal, it is implausible to claim that the resulting actions or values do not reflect my autonomy. So since we can imagine *any* such preference as being autonomously formed, given a fantastic enough situation, then it cannot be the *content* of the preference that determines its autonomy. It is always the *origin* of desires that matters in judgments about autonomy.

On this view, Christman argues, as long as an individual's desires and values are generated in accordance with the procedural conditions of autonomous preference formation, then the actions that stem from them will be positively free, regardless of the *content* of those desires and values.

Three comments are in order regarding Christman's account of positive liberty. First, his conception of positive liberty is both similar to and different from conceptions of autonomy in medical ethics (e.g., Faden and Beauchamp's theory). Christman's account is different from medical autonomy in at least one important respect. By focusing on the autonomous generation of preferences and desires, it broadens the scope by which particular actions can be judged autonomous or non-autonomous. Under Christman's view of positive freedom, a Jehovah's Witness who was indoctrinated as a young child or whose beliefs were oppressively imposed on her would be positively unfree, no matter how no matter how genuinely the beliefs and desires are felt. If she finds herself in a position where she needs a blood transfusion, she would not be able to autonomously refuse one insofar as her refusal is based on beliefs that were not generated autonomously.[6] Whether or not this view is appropriate, Faden and Beauchamp avoid it by focusing narrowly on autonomous *actions* rather than autonomous *persons* or autonomously generated *desires*. Under their construal of autonomy, the procedures by which one's desires are formed matter less than the procedures by which the goals of those desires are attained. When the woman finds herself in the position to decide on whether to agree to a blood transfusion, the specific desires that emanate from her

---

[6] There seems to be a possibility, under Christman's account, for one to come to autonomously possess beliefs and desires that were imposed on her through indoctrination. For instance, if the Jehovah's Witness were to critically reflect on and accept the beliefs and desires, perhaps later on in life, then they would be autonomous. The example above assumes she has not done this.

religious beliefs are scrutinized only to the extent that they produce a decision that is made *intentionally* (i.e., in accordance with a mental blueprint for action) and with sufficient *understanding* of the nature of the action and its foreseeable consequences (Faden and Beauchamp 1986, 242-252). She could, therefore, autonomously refuse a blood transfusion, under Faden and Beauchamp's account, based on ingrained beliefs and desires that were not generated autonomously under Christman's account. Conversely, she could possess autonomously-generated desires but still be unable to autonomously refuse a blood transfusion if she cannot satisfy the conditions of intentionality and understanding.

Despite the major difference described above, there is an essential commonality between Christman's and Faden and Beauchamp's notions of autonomy. They both take a neutral position either on the person's choices and actions or the content of the desires from which they descend. While Christman's version of autonomy is explicitly neutral with regard to the content of an individual's desires, Faden and Beauchamp's theory of autonomy takes a neutral stance on the individual's actual decisions or actions. Like the conditions for the autonomous generation of desires, the conditions of *intentionality* and *understanding* (the agency conditions) are procedural conditions whose satisfaction is independent of whether the patient makes the so-called *right* decision. As long as the individual decides on a course of action intentionally and with sufficient understanding, whether or not the decision accords with the recommendation of the physician, the decision is autonomous. This impartiality is fundamental to patients' rights to decide for themselves.

The second remark on Christman's notion of positive liberty pertains to his claim that his procedural conditions are immune from the paradox of positive liberty (i.e., the inversion thesis) articulated by Berlin. This claim rests on the view that it would be practically impossible for anyone to have sufficient knowledge about another's internal desires to evaluate whether they are rationally consistent and transitive. This may be true for the general sphere of political and social liberty that Christman seems to be addressing, but it would be mistaken to assume that it also applies to the medical sphere. Evaluations of medical decision-making capacity are specifically designed to determine whether a patient is able to autonomously make decisions about his or her medical treatment. Though a clinician's capacity determination would not involve an evaluation of a patient's decision against Christman's criteria, the close scrutiny of the reasons behind a patient's decision, which is what a capacity assessment provides, would certainly make that possible. In fact, contemporary standards for decision-making capacity are based on a particular normative conception of autonomy, and there is no reason Christman's conception could not replace it as the basis for determining capacity. As such, regardless of which specific version of autonomy underlies capacity determinations, even if it is a content-neutral version, the basis for coercive treatment (i.e., a hindrance of negative liberty) is a person's positively unfree decision. In this sense, Christman's view of positive freedom, although perhaps more benign than the more nefarious versions Berlin warned against, retains the logic of positive freedom articulated in Berlin's inversion thesis.

The third element of Christman's argument that needs to be addressed is perhaps the most important. He claims that his version of positive freedom, at least the external evidence condition, accords with the views of most adherents to negative liberty. If interference with one's "unwitting actions" (Christman, Liberalism and Individual Positive Freedom 1991, 355) does not constitute a meaningful disruption of his or her self-government, as, he claims, even adherents to negative liberty accept, then the use of some minimal version of rationality as a precondition for negative liberty is acceptable. Even John Stuart Mill's notion of liberty is predicated upon some form of rationality:

> If either a public officer or any one else saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back without any real infringement of his liberty; for liberty consists in doing what one desires, and he does not desire to fall into the river. Nevertheless, when there is not a certainty, but only a danger of mischief, no one but the person himself can judge the sufficiency of the motive which may prompt him to incur the risk: in this case, therefore, (unless he is a child, or delirious, or in some state of excitement or absorption incompatible with the full use of the reflecting faculty,) he ought, I conceive, to be only warned of the danger; not forcibly prevented from exposing himself to it (Mill 1989, 142).

The justification Mill provides for preventing a person from unknowingly crossing an unsafe bridge (when there is no time to warn him of the danger) is that it is not a real infringement upon his liberty because liberty consists in doing what one desires, and the person does not desire to fall into the river. Of course, Mill presumes, reasonably, that the person does not want to fall in the river. He seems to leave room for this presumption to be rebutted, however. For seizing the man to be justified, there must be no time to warn him. Perhaps if the man were warned beforehand and decided to cross the bridge anyway, Mill would find interference unjustified. Or perhaps even if the man were seized and insisted on continuing over the bridge, further interference with his actions would be unjustified. If interference after warning or temporary seizure is justified, however, then Mill's idea of liberty contains at least a kernel of positive liberty which endorses certain values. If the man is not free to knowingly walk over the unsafe bridge, an act that will result in harm to him, then Mill's notion of liberty must contain some value preconditions. Regardless of which side Mill would take, his primary point is that it is justifiable to disrupt one's actions if the actions do not comport with what the person desires.

Mill's position on cases in which the outcome of an action is uncertain is clearer. Only the actor, he argues, can decide whether or not to undertake an action that carries with it the risk of harm. In such cases, the person should not be forcibly prevented from incurring the risk, but he should be warned of it. Of course, this realm of liberty only extends to those who are rational. Children, the delirious, and persons with temporary impediments to their rational faculties may be justifiably coerced. This seems to support the contemporary structure of autonomy in medical ethics. If negative liberty is only extended to those who have some form of rationality or autonomy, then those who lack such rationality or autonomy do not have a right against being compelled against their will. But Mill is only disqualifying from the right to liberty those irrational or non-autonomous individuals who wish to engage in activities that are harmful to themselves, or at least potentially harmful to themselves. Some minimal form of autonomy or rationality, then, seems to be a precondition for the right to engage in harmful or potentially harmful activities without interference (negative liberty). This principle, though, does not disqualify irrational or non-autonomous individuals from acting on *all* of their empirical wishes, as Faden and Beauchamp's autonomy as a gatekeeper concept does. Rather, harm to one's self is the threshold beyond which non-autonomous individuals' right to liberty is justifiably hindered. What, then, is the source of a non-autonomous individual’s general right against coercion? For sure, it is possible that Mill or others would not even regard a non-rational person as an appropriate bearer of liberty rights. Though Mill’s passage above does not disqualify non-autonomous individuals from acting on *all* of their wishes, it does not guarantee their right to act on non-harm-producing wishes either. As we saw in chapter two, however, there are instances in which non-autonomous individual’s certainly have a right against being coerced into activities against their wishes.

Richard Fallon’s Concept of Ascriptive Autonomy

Even if there are justifiable reasons for coercing individuals who do not have the capacity to act autonomously, this does not necessarily mean that the empirical wishes of those individuals do not matter, that they can be pushed aside. I would argue that the wishes of non-autonomous individuals are morally significant, and that the lack of autonomy, in itself, does not disqualify them from making their own decisions and acting on them. Richard Fallon’s notion of ascriptive autonomy, “the autonomy we ascribe to ourselves and others as the foundation of a right to make self-regarding decisions” (Fallon 1994, 878), helps to illuminate the moral significance of the wishes of all people, even non-rational agents. Ascriptive autonomy is different from *descriptive autonomy*, which “refers to people’s actual condition and signifies the extent to which they are meaningfully ‘self-governed’ in a universe shaped by causal forces” (Fallon 1994, 877). Although Fallon attributes the idea of ascriptive autonomy to Kant and points out that autonomy in the Kantian tradition presupposes a freedom of the will, he is hesitant to suggest that ascriptive autonomy must be premised on such a notion. If it were established that no individuals have freedom of the will, he argues, our lives would not likely be different because we would still resent paternalistic actions and we cannot help but experience ourselves as bearers of free will. That, in itself, he argues is sufficient to establish ascriptive autonomy as an important moral and political value – “our experience of ourselves as moral agents with both the capacity and the right to make decisions for ourselves, even when those decisions are insufficiently informed, self-aware, and self-critical to count as autonomous under any very stringent standards of descriptive autonomy” (Fallon 1994, 893).

There is an intuitive appeal to Fallon’s idea. An individual who is forced against his will to do something, regardless of whether the interference is justified according to some moral standard, is likely going to feel wronged. If there is value in the mere experience of ourselves as being both able to and having the right to make our own decisions, whether or not those decisions stem from some form of rational deliberation, then it is inadequate to base one’s right to negative liberty solely on his ability to decide and/or act rationally or autonomously. By also grounding negative liberty in the right people ascribe to themselves to make their own decisions, we can begin to see the moral significance of even their irrational wishes and desires. By doing so, the right to decide not to engage in sexual relations, participate in research, or have an abortion stands on its own, and it is not dependent upon one’s capacity to autonomously refuse.

This, of course, is not to say that a non-autonomous refusal cannot be overridden. Rather, it is to highlight that even when a non-autonomous refusal is overridden, there remains a tension between the moral significance of the individual’s empirical wishes and the moral right (or even obligation) to override them. This conflicts with the contemporary structure of autonomy. If one’s empirical wishes (even the non-autonomous ones) are morally significant and bear on one's right to negative liberty, then medical agency (or decision-making capacity) cannot be a necessary condition for the right to (negative) liberty, at least not in all instances. Decision-making capacity, then, does not rightfully serve as an all-encompassing gatekeeper to one's right to make his own decisions. If it did, then a non-autonomous individual would not have the right to make *any* decisions. If he does not have the right to make any decisions, then he is merely *permitted* by someone else to make certain decisions, perhaps because the decisions are viewed as inconsequential by the other person, but not because there is any moral significance to his wishes. Denying the moral significance of descriptively non-autonomous decisions cheapens those decisions, and it diminishes the moral worth of those who make them.

If the empirical wishes of non-autonomous individuals are morally significant, then the question is at what point, or under what conditions, does a non-autonomous individual cease to have the right to make a decision? Or, under what conditions is it acceptable to override a non-autonomous individual's wishes? This shifts the emphasis from the appropriateness of a particular standard for surrogate decision making to the question of whether the surrogate decision maker even has the moral right or authority to decision on the patient's behalf. It adds an intermediate step between a capacity determination and a surrogate's overriding decision (if appropriate).

CHAPTER 4

In the preceding chapters, I argued that a non-autonomous refusal is morally significant and that the lack of decision-making capacity should not automatically deny an individual the right to have his refusal respected. I argued that a person's negative liberty rights are not only grounded in his ability to act rationally but also in his experience of himself as having the ability and right to make his own decisions. If this is true, then the notion that medical agency is a necessary condition is false and requires modification. Of course, my intent is not to insist that non-autonomous patients have an unfettered right to refuse any medical treatment they do not wish to undergo. Rather, my intent is to challenge the idea that forcing medical treatment on an unwilling person is justified simply because it does not violate his autonomy. A non-autonomous person is not merely a vacant vessel whose empirical wishes can be cast as irrelevant just because he is not autonomous. The individual's empirical wishes should be taken into account, but they should not be decisive in all situations. In this chapter, I will propose some conditions that, if satisfied, justify not abiding by an individual's refusal of medical treatment.

Paternalism

Although I am only proposing conditions for not abiding by a patient's non-autonomous refusal, and not conditions for coercively treating him, the effect of not abiding by his refusal could very well be paternalistic action if the surrogate consents to the treatment. Thus, it is fitting to develop these conditions with justifications for paternalistic treatment in mind. Gerald Dworkin defines paternalism as "roughly the interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced" (Dworkin 1972). Beauchamp and Childress define it as "the intentional overriding of one person's preferences or actions by another person, where the person who overrides justifies this action by appeal to the goal of benefiting or of preventing or mitigating harm to the person whose preferences or actions are overridden" (Beauchamp and Childress 2009, 208). Paternalism can also be broken down into two distinct concepts: *hard* paternalism and *soft* paternalism. Hard paternalism is when the subject of the paternalistic action is autonomous. The individual is able to autonomously pursue a course of action, and accept the harms or risks involved, but he is not allowed to do so by the person or entity acting paternalistically. Soft paternalism is when the subject of the paternalistic action is not autonomous. He is unable to autonomously pursue a particular course of action, so he is treated paternalistically to either protect him from harm or secure him some benefit (Beauchamp and Childress 2009, 210). The concept of soft paternalism is what is under consideration here.

Person-Mediating Maxim versus Act-Mediating Maxim

Alan Sloble points out that a principle of paternalism can be based on one of two logically distinct maxims: a person-mediating maxim or an act-mediating maxim. A person-mediating maxim refers solely to the properties or characteristics of the individual who is subjected to the paternalistic intervention. According to a person-mediating maxim, "we are justified in intervening only when the person has (or lacks) certain properties, no matter what act or acts she intends to perform" (Sloble 1982, 5). For example, a person who lacks decision-making capacity is non-autonomous, so we paternalistically intervene and force him to undergo needed treatment. This is the maxim upon which the notion of “soft paternalism” is based, and it describes the current justification for treating patients against their expressed wishes.

An act-mediating maxim refers solely to the properties of the acts that are subject to prohibition. According to an act-mediating maxim, “we are justified in intervening only to prevent certain acts or activities, no matter who wants to perform the act and no matter what her state of mind happens to be” (Sloble 1982, 5). This could be applied to medical treatment as follows: the property of forgoing needed treatment leads to negative outcomes for the individual who refuses it, so we paternalistically intervene and compel treatment, regardless of whether or not he is autonomous. As noted above, determinations of whether to intervene in treatment decisions is made according to a person-mediating maxim. The properties of the act of refusal itself are not considered. The doctrine of informed consent implies a prima facie right to autonomously refuse *any* treatment. The properties of the act of drug use, however, are referred to as justifiable reasons for prohibiting all individuals from using drugs, and no regard is given to the properties of the individual drug users themselves. It does not matter, for instance, if an individual is able to autonomously choose to use cocaine. Therefore, drug prohibition reflects an act-mediating maxim.

Paternalism in the Case of Non-Autonomous Active Refusal

To account for the moral significance of a non-autonomous individual’s active refusal, but still not allow this moral significance to function as an absolute determinant, I propose that a principle for paternalistic intervention incorporate both a person-mediating maxim and an act-mediating maxim. The person-mediating maxim would refer to whether or not the individual possesses the requisite characteristics of autonomous medical decision-making. For the purposes of this analysis, I will not attempt to develop an original set of properties or characteristics a person needs in order to autonomously refuse medical treatment. Instead, I will defer to contemporary notions of autonomy in medical ethics, namely Faden and Beauchamp’s criteria that a refusal be intentional, with sufficient understanding, and in substantial absence of controlling influences (Faden and Beauchamp 1986, 238). If the individual possesses the requisite characteristics of autonomy by satisfying these criteria, then his active refusal is binding on the physician, provided there are no other reasons (e.g., the possibility of transmitting an infectious disease to others[7]) to override his decision.

If a principle for paternalism in the case of a non-autonomous refusal is going to incorporate both a person-mediating and an act-mediating maxim, then the lack of autonomy alone is not sufficient justification for overriding a patient’s refusal. Before proposing an act-mediating maxim for evaluating a refusal, one point must be addressed. One might argue that decisions to override active refusals of non-autonomous patients already use both a person-mediating maxim and an act-mediating maxim. This is probably true, for the concept of paternalism involves coercing people for their own good. If the justification for a particular instance of coercion (or a rule for coercive action) does not refer to the characteristics of the act being prohibited or compelled, then it is not appropriately termed paternalism. Furthermore, justifying the compelled chemotherapy treatment of a non-autonomous person against his active refusal without reference to the potential consequences of the refusal (e.g., uncontrollable progression of disease, which could potentially lead to death) hardly seems satisfactory. If an appeal to

---

[7] Overriding the autonomous person’s refusal of a treatment or taking another course of action against her wishes (e.g., quarantine) would be paternalism in accordance with an act-mediating maxim.

the potential consequences of the refusal is unnecessary, and referring solely to the individual’s status as non-autonomous is sufficient, then one could presumably justify coercive treatment to achieve any number of ends, not just those which are perceived to benefit the patient. Forcing a non-autonomous individual to undergo chemotherapy that does not provide the prospect of medical benefit, however, is not acceptable, so, one could argue, there must be at least an implicit recognition that the properties of the specific refusal are relevant.

I do not dispute that there are limits to a physician’s moral authority to coerce a non-autonomous patient, but I would hypothesize that the limits, in practice, could also be simply a reflection of the physician’s duty to provide potentially beneficial treatment to patients. For example, a physician would be violating her professional duty if she proposed a treatment that was unnecessary, futile, and/or unacceptably harmful, so that duty, by itself, restricts the types of action the physician may propose to those conforming to professional standards. Because the treatments proposed by a physician are, at least theoretically, limited to those offering potential medical benefit according to her clinical judgment (but within boundaries of professional standards), this creates a presumption in favor of the physician’s proposed course of action. Decisions against a physician’s proposed treatment are not the *right* decisions (from a medical standpoint) insofar as they conflict with the proposal that offers the best chance of *medical benefit*. Autonomous individuals, though, may refuse to consent to a proposed treatment because they need not value medical benefit to the same extent as the medical profession, or even at all. Non-autonomous individuals are not afforded this luxury. If there is an act-mediating maxim operating when a patient’s non-autonomous refusal is overridden, it is difficult to distinguish it from the notion that they may not deviate from a physician's medical advice and the surrogate decision maker's consent, which grants the physician the moral authority to treat in accordance with that advice.

Regardless of whether there is an act-mediating maxim operating implicitly when non-autonomous refusals are overridden, such a maxim is still not part of the explicit theoretical justification for paternalistic treatment. The *autonomy as a gatekeeper concept* and the resultant structure of autonomy in medical ethics nullify the right of individuals to make *any* decision they cannot make autonomously. Buchanan and Brock's *sliding scale* notion of decision-making capacity does consider the properties of the act of refusing (or consenting to) medical treatment, but only by intertwining them with the properties of decision-making capacity (and, thus, autonomy) itself. According to the sliding scale concept, "competence is a *relational property* determined by a *variable standard*. Whether a person is competent to make a given decision depends not only upon that person's own capacities but also upon certain features of the decision itself – including risk and information requirements" (Buchanan and Brock 1989, 60). The standards for decision-making capacity can, therefore, be raised or lowered depending on the nature of the decision the individual faces. This makes relevant both the properties of a particular person and the properties of a particular decision (the act). As Buchanan and Brock point out, "you may decide that your five-year-old child is competent to choose between a hamburger and a hotdog for lunch, but you would not think the child competent to make a decision about how to invest a large sum of money" (Buchanan and Brock 1989, 60). According to the authors, this "is because the risk in the latter case is greater, and the information required for reasoning about the relevant consequences of the options is much more complex" (Buchanan and Brock 1989, 60). The shortcoming of this concept, besides the mistaken assumption that risky decisions are more complex and, thus, require a higher level of capacity than non-risky ones (Beauchamp and Childress 2009, 116-17), is that the requisite properties of the person are dependent upon the particular properties of the act. This is probably true, but it is not helpful for determining the limits of a non-autonomous patient's right to refuse unwanted treatment. There is really only one maxim here, and it is a person-mediating maxim. In determining whether paternalistic treatment is justified, it refers only to the properties of the person (i.e., whether or not the person has capacity), even if the appropriate properties of the person are determined by the nature of the decision (i.e., the act).

In order to acknowledge the moral significance of a non-autonomous individual's refusal, the person-mediating maxim and the act-mediating maxim must be independent of one another. For treatment to be coerced, one must determine both that the individual is not able to autonomously refuse the proposed treatment (he lacks decision-making capacity) and that the properties of the refusal are such that treatment against his expressed wishes could be justified. By requiring decisions to reject a non-autonomous refusal to be made against both maxims, the doctrine of informed consent (and informed refusal), and the significance of autonomy upon which it is based, is preserved. If the person is autonomous, his refusal is authoritative, and it may not be overridden, regardless of the consequences of the act of refusal itself. If the person is non-autonomous, he is not automatically exposed to the possibility of his wishes being overridden. Instead, the properties of the act of refusal must be examined first. If they cross a threshold beyond which coercion is justified, then the individual's refusal of treatment is not authoritative. What are these properties, and what is the appropriate threshold for limiting a non-autonomous person's right to refuse?

Harm Principle

A good place to start is with the approach taken by Diekema in his argument against a "best interests" standard for justifying state intervention in cases where parents refuse medical treatment for their children. For various reasons, he argues that a harm principle is more suitable. For one, it may be difficult to precisely define the "best interest" of a child. The standard is most straightforwardly applied in situations in which a child's life is threatened and death can be avoided by administering safe and effective treatment. Less serious threats to a child's health (e.g., cleft lip and palate repair), however, may prove more difficult to determine whether a parent's refusal violates the standard (D. S. Diekema 2004). This is true in the case of non-autonomous refusals as well. Where a patient's life is in danger, and that danger can be averted with a very safe and effective treatment, little controversy would surround a decision not to abide by the patient's refusal. It is less clear, though, whether a burdensome treatment with severe side effects and only a moderate chance of success is truly in a patient's best interests. Furthermore, it is even less clear whether the provision of a treatment that is not necessary for survival, but is very risky and/or carries burdensome side effects is truly in the best interests of the patient. Even if the treatment is not burdensome and is conceivably in a patient's best interest, it does not always seem appropriate to disregard his refusal just because it is not autonomous. For example, a patient may be in excruciating back pain that disturbs his daily functioning. Treatment with narcotics would alleviate the pain, but the patient fervently opposes such treatment. However, he lacks decision-making capacity, so he is unable to provide an informed refusal. The relief of the oppressive symptoms is likely in the patient’s best interests, but it does not seem defensible to simply discount his vehement refusal.

This highlights the complexity of interests at play in medical decision making. As Diekema points out, medical professionals frequently reduce best interests to objective medical interests. For instance, in discussing chemotherapy for a child with cancer, physicians emphasize the increased chance of survival offered by the treatment while underestimating its adverse effects. Some parents, on the other hand, may judge that the increased chance of survival does not outweigh the risks, side effects, discomfort, and other life disruptions that accompany the treatment. Reconciling these conflicting perspectives is no easy task (D. S. Diekema 2004).

For the reasons discussed above, a best interest standard is not the best threshold for determining the limits of a non-autonomous refusal. Therefore, I will propose a harm threshold similar to that offered by Diekema. It should be noted at the outset that these are not analogous scenarios. Diekema’s article concerns the authority of the state, under the doctrine of *parens patriae*, to supersede an individual’s parental authority over his or her child, specifically his or her authority to refuse appropriate medical treatment (D. S. Diekema 2004). The parents in Diekema’s analysis are presumed to be autonomous, and it is the effects (i.e., the harm) their decisions have on their children that are at issue. The matter under consideration in this analysis is the moral authority of a surrogate decision maker, which could include the state, to supplant a non-autonomous individual’s expressed refusal of appropriate medical treatment. It is the effects of the patient’s refusal (i.e., the harm that could result) that trigger the surrogate’s moral authority. Despite the differences between the two scenarios, the act-mediating maxim proposed by Diekema can be adopted for non-autonomous refusals.

Diekema notes that not *all* harm should trigger state interference. As such, he proposes eight conditions for justified state interference with parental decision-making. First, the parental refusal to consent to treatment will place the child at significant risk of serious preventable harm, and some other course of action must be available that would prevent the harm to which the child is being exposed. Diekema admits that determining whether the harm is *significant* and *serious* enough to justify intervention invites subjectivity, but, he maintains, this is inescapable, and there "will always be an interpretive element in determining whether a parental decision crosses the threshold for state intervention" (D. S. Diekema 2004, 253).

Second, the harm must be imminent and require immediate action to prevent it. If a parental refusal does not place a child at imminent risk of significant and serious harm, state intervention should be postponed and attempts should be made to resolve the issue with the child's parents. Third, the treatment plan refused by the parents should be proven effective. There should be consensus among experts that interference with the refusal and the provision of treatment would likely "prevent, eliminate, or reduce the harm in question" (D. S. Diekema 2004, 253).

Fourth, overriding the parent's refusal and providing treatment must be necessary to prevent the serious and significant harm that the child faces. If there are any acceptable alternatives that are less intrusive to parental autonomy, one of these ought to be pursued in favor of state intervention. For intervention to be justifiable, "all alternatives to interference with parental decision-making must have been explored and no morally preferable alternative found to exist” (D. S. Diekema 2004, 254). Fifth, a net benefit to the child must result from the interference with the parental refusal. The harm prevented must be greater than the potential harms of the treatment and the harm that could result to the integrity of the family. The intervention “must be clearly preferable to the course of action proposed by the parents, and its projected benefits must outweigh its burdens” (D. S. Diekema 2004, 254).

Sixth, the extent of the state’s intervention and the treatment provided under the state’s authority should be the least intrusive alternative that will reduce the harm to the child and minimize the impact on parental authority. In other words, the treatment over the parent’s objections “should include only that which is necessary to prevent the harm to the child” (D. S. Diekema 2004, 254). Seventh, the state intervention must be impartial and generalizable to all similar cases. Morally irrelevant considerations, such as the religious nature of a decision, should not influence the decision. Diekema uses state intervention in a parental refusal of blood transfusion as an example. He argues that intervention is not justified because of the religious nature of the parental refusal, but rather because the parents are refusing a potentially life-saving therapy that meets all the other conditions. Therefore, the likely outcome of the refusal, that it will result in serious harm to the child, and not the reasons for the refusal, is what is important. The final condition is that the state intervention must pass the test of publicity; that is, other parents would agree that the outcome of state intervention is appropriate for all children (D. S. Diekema 2004).

The conditions provided by Diekema can be adopted and modified to produce a set of conditions for determining whether a non-autonomous patient’s refusal of treatment should be respected. It is important to reiterate that the following conditions are not intended as a model for surrogate decision-making. Rather they represent an intermediate step between determining incapacity and seeking a decision from a surrogate decision maker (provided the conditions are met). In essence, the conditions are for determining whether the patient's non-autonomous refusal should be substituted with the decision of a surrogate. As such, they are not even intended to provide guidance on whether coerced treatment is justified. Because it is only an intermediate step that does not bind a surrogate to a particular, it is possible for her to refuse the treatment on the patient's behalf. In such a case, the outcome of the surrogate decision-making process adheres to the patient's expressed wishes, the physician has no moral authority to treat, and no coercion actually takes place.[8] If the surrogate consents to the treatment on the patient's behalf, however, the physician has the moral authority to provide the treatment against the patient's wishes. If the patient's refusal does meet the following conditions, then the moral significance of his non-autonomous refusal is sufficient for his expressed wishes to be respected, and a surrogate does not have the right to make a decision on his behalf.

The first condition, mirroring offered by Diekema, is that the patient's refusal will place him or her at significant risk of serious preventable harm. As in the case of overriding a parent's refusal of treatment for his her child, the terms *significant risk* and *serious* harm are open to interpretation. Nonetheless, such qualifiers are necessary to distinguish the appropriate level of harm from more trivial harms, such as moderate pain or discomfort, or harms that are possible but very unlikely to occur. Furthermore, it is

---

[8] Although, by definition, the surrogate's refusal would be harmful to the patient (if patient's refusal crosses the harm threshold, the surrogate's refusal on his behalf does also), this leaves room for the surrogate to refuse life-saving or life-sustaining treatment in accordance with the patient's previously expressed wishes.

nearly impossible to put forth a precise level of risk and seriousness of harm that will be suitable for all cases, so it is appropriate to allow for some flexibility.

Second, the harm must be imminent and require immediate action to prevent it. If this is not the case, then the patient's refusal should be respected, though attempts could be made to persuade him to undergo the proposed treatment. Third, the treatment proposed by the physician should be proven effective and have a high likelihood of successfully preventing, eliminating, or substantially reducing the level or risk of harm with which the patient is confronted. Diekema did not explicitly specify a level at which harm reduction is satisfactory, but it is necessary to distinguish between very small reductions in harm and *substantial* reductions that have a meaningful impact on the overall level of harm the patient faces. Although the term *substantial* is also open to interpretation, it can be given a certain degree of precision. For the purposes of this condition, harm is *substantially* reduced if the harm the patient faces no longer qualifies as *serious* harm, or if the risk of harm is no longer *significant*.

Fourth, the provision of the proposed treatment must be necessary to prevent, eliminate, or substantially reduce the level or risk of harm. Diekema proposes that if there are acceptable alternatives that are less intrusive to parental autonomy, then these should be pursued instead. For non-autonomous patients refusing a specific medical treatment, if there are available alternatives with an equal or similar likelihood of success, and the patient agrees to undergo one of those, then there is no need to determine whether he has the right to refuse any specific treatment. Instead, the approach with which he agrees should be pursued since there are no discernible differences (in terms of harm aversion) between it and the initially proposed treatment. If he refuses *any* treatment, however, then the evaluation should include whether at least one of the available forms of medical treatment is necessary.

Fifth, the likelihood and magnitude of harm prevented, eliminated, or reduced must be greater than the likelihood and magnitude of the harm from both the treatment and the coercion. One might argue that this condition should exclude from consideration the potential harm from the coercion itself. Since these conditions represent an intermediate step, the objection might go, what is at issue is whether the non-autonomous refusal should be respected, not whether the person should be forcibly treated (which would be a question the surrogate would answer). I would respond by pointing out that the intent of this intermediate step is to acknowledge the moral significance of the patient’s refusal, and since a determination not to accept the refusal could potentially lead to coerced treatment, the experience of harm via the coercion itself is relevant.

Sixth, only the likely outcomes of the refusal should be considered. The reasons the patient gives for his refusal should not. At this point in the process, it has already been determined that the individual is not able to autonomously refuse the proposed treatment. Thus, his reasons for refusing the treatment have already been deemed inadequate. Nonetheless, the reasons are irrelevant to whether the refusal would place the patient at significant risk of serious and imminent harm that could be prevented, eliminated, or substantially reduced by the treatment.

Diekema’s sixth condition, that the extent of the intervention and the treatment be the least intrusive alternative, was excluded from the conditions offered above. This condition is intended to constrain the extent of the intrusion to that which is necessary for preventing the harm. Since the fourth condition requires that the treatment be necessary, it is not clear this condition adds anything else. Diekema's eighth condition was also excluded. While it may be important to anticipate the reaction of the public, this is not necessary for determining whether or not the patient faces significant and serious harm if he does not undergo a particular treatment.

The Case of Ryan

This section will apply the six conditions to the case of Ryan, which was introduced in the first chapter. It will be assumed that he lacks decision-making capacity and cannot provide an informed refusal, so the emphasis will be on the act-mediating maxim. The application is intended only to place Ryan's situation within the framework of the harm principle; it is not intended to reach a conclusion as to whether his expressed refusal of treatment should be respected.

The first question is whether Ryan's refusal of antipsychotic treatment will place him at significant risk of serious preventable harm. Ryan appears to be having a fairly severe psychotic episode. The primary harms that could result are from behaviors that Ryan might engage in as a result of his condition. The potential harm, therefore, is less predictable than, say, the harm that could result from a rapidly growing tumor on his spleen. If the physician determines that without treatment Ryan is likely to commit suicide, then Ryan's refusal of treatment places him at significant risk of serious harm. This is a possibility. Suicide rates in individuals with schizophrenia are higher than in the general population, and attempts occur most frequently in the early course of the illness (Palmer, Pankratz and Bostwick 2005). Additionally, hallucinations and delusions are among the illness-related predictors of suicidality in schizophrenic patients (Hor and Taylor 2010). If Ryan is not likely to commit suicide or otherwise inflict serious harm upon himself, but, instead, his symptoms could lead to certain social harms, such as having to drop out of college, then his refusal does not place him at risk of serious harm.[9] It could also be suggested that the symptoms themselves constitute serious harm to Ryan. This is an appealing suggestion. If one were to consider another scenario, in which there was a risk that someone would develop the symptoms from which Ryan is currently suffering, then it is would be reasonable to describe the potential harm as *serious*. Even so, the harm has already occurred to Ryan, so he is not at *risk* of serious harm. For this reason, the symptoms of his condition should not be included in the calculation.

The second question is whether the harm faced by Ryan is imminent and requires immediate action to prevent it. Since predicting behavior is an imprecise exercise, it makes determining the imminence of a particular form of harm such as suicide or other self-inflicted bodily harm all the more difficult. If there was reasonable certainty, according to the physician's clinical judgment, that Ryan would commit suicide soon after leaving the hospital, then this could reasonably be deemed an imminent harm. Such a clinical prediction would not be implausible, given that his hallucinations and delusions are predictive of suicidality (Hor and Taylor 2010) and the fact that suicide attempts occur most often at the beginning of disease onset (Palmer, Pankratz and Bostwick 2005). If it is merely predicted that Ryan is likely to commit suicide in his lifetime, then this could not justifiably be described as imminent harm.

The third consideration regards the antipsychotic therapy that the physician would like to use to treat Ryan. If Ryan is at significant risk of serious and imminent harm, but there is no effective method of protecting him against that harm, then his refusal should

---

[9] Of course, if Ryan poses a risk of harm to others, then coerced treatment or confinement can be justified on those grounds. This, however, is not the purpose of this analysis, which is intended to specify the level of harm to oneself that justifies not respecting one's refusal of treatment.

be respected. Suppose the antipsychotic agents would relieve some of his symptoms (providing some benefit), but there would still be no assurance that he would not seriously harm himself. This condition would not be satisfied because the issue is whether the therapy protects from the harm, not whether it provides him any benefit. There is evidence that certain antipsychotic agents do reduce the risk of suicide, however, (Meltzer, et al. 2003), so it is feasible that this condition can be satisfied, provided the treatment is likely to be effective in Ryan's particular case.

The fourth question is whether the provision of antipsychotic treatment is necessary to prevent, eliminate, or substantially reduce the level or risk of harm. If the anticipated harm is suicide or some other self-inflicted bodily harm, then there are non-treatment options, such as confinement in the psychiatric ward (itself a deprivation of his liberty) or close supervision by others (e.g., Ryan's family or friends). Of course, whether non-professional supervision is a realistic option depends upon a number of factors, such as the ability and willingness of the relevant individuals to take on the task. If Ryan's parents are unable to devote sufficient time to supervise his behavior, or if they are unwilling to, perhaps insisting that he needs professional supervision, then this is not a viable alternative. Depending on the particular circumstances, a case could be made that the treatment is necessary to protect Ryan against the harms associated with his refusal.

The fifth question is whether the likelihood and magnitude of the harm prevented, eliminated, or reduced by antipsychotic treatment would be greater than the likelihood and magnitude of the harm from both the treatment and the coercion. A significant proportion of schizophrenic patients fail to adhere to their antipsychotic medication regimen. This is largely due to the undesirable side effects of antipsychotic medications, which lead to durable negative attitudes toward antipsychotic therapy in general (Lambert, et al. 2004). Though the side effects and the negative attitudes toward treatment that follow certainly qualify as harms, they are not as serious as death from suicide. As such, provided the other conditions are met, it is unlikely that the harms associated with antipsychotic treatment will outweigh the harm averted by not deferring to Ryan's refusal.

The sixth condition is more for conceptual guidance than for determining whether the refusal will lead to serious preventable harm. It is intended mainly to ensure an impartial assessment of the likely harms of the refusal. In Ryan's case, however, it might be difficult to draw a hard line between evidence for the predicted outcomes and an evaluative stance on the reasons Ryan gives for refusing treatment. For instance, Ryan does not want to undergo antipsychotic treatment because he believes the physician is a CIA agent who intends to kill him with the drugs. This is both Ryan's expressed reason for refusing treatment and an indicator (e.g., delusion) of the potential harm that could result from his refusal. Still, a conceptual distinction can be drawn. Although Ryan's reason for refusing treatment is irrational and evidences lack of decision-making capacity, it should only be taken into consideration to the extent that it predicts the harm that could result from not receiving treatment. If Ryan's expressed reasons are indicative of potential harm, then they are relevant factors for determining whether the other conditions are met.

As the application of the harm principle to Ryan's case shows, it is possible to appreciate the moral significance of a non-autonomous patient's refusal of beneficial treatment without granting that patient an absolute right to refuse. If the particular circumstances indicate that Ryan’s refusal will lead to serious harm that can be averted only by medical treatment, then it is acceptable to protect him from his expressed wishes and not concede to his refusal. Of course, the fact that it is justifiable to override Ryan’s wishes does not diminish the moral significance of his refusal. Rather, there is still tension between this moral significance and the obligation to protect him from the consequences of his decisions. He does not have the capacity to autonomously accept the consequences of his refusal decision, but this does not mean that he is just an autonomy vacuum whose empirical wishes and desires are irrelevant to what is done to his body. His wishes and desires are relevant, but the enormity of the harm that will result from them could outweigh their moral significance.

Conclusion

The idea that there is a moral tension between a non-autonomous patient’s refusal and a decision to treat him against his wishes does not seem that controversial. It is unlikely that one could coerce another human being, autonomous or not, without sensing this tension, regardless of whether good intentions were the basis for the coercive act. Consider the intuitive moral discomfort that must accompany strapping down a non-autonomous patient and forcing him to undergo an unwanted treatment. The contemporary justification for coercing non-autonomous patients does not account for this discomfort. Its silence on it effectively denies the moral significance of the non-autonomous person’s empirical wishes. The failure to acknowledge this moral significance allows the forcible treatment of patients to be justified by simply referring to their lack of autonomy.

The purpose of this thesis was to insert the moral significance of a non-autonomous individual's wishes into the theoretical justification for coercing him. In doing so, I introduced some constraints on the moral authority of surrogates to authorize the forcible treatment of patients who lack decision-making capacity. Rather than moving directly from a determination of incapacity to considerations about the right treatment decision, one first must determine whether the consequences of the patient's refusal were such that it should not be respected on its own. Although I proposed the use of a harm principle to make such a determination, my overarching argument that a patient's lack of decision-making capacity does not automatically justify coercing him does not rely on the appropriateness of the specified principle as a constraint. While I believe harm is the appropriate principle for balancing the moral significance of a patient's refusal and the importance of protecting him from his decisions, my argument does not preclude the use of another formulation as a constraint on a surrogate's moral authority.

BIBLIOGRAPHY

Alderson, Priscilla. "Competent Children? Minors' consent to health care treatment and research." *Social Science and Medicine* 65 (2007): 2272-2283.

Alzheimer's Association. "Research Consent for Cognitively Impaired Adults: Recommendations for Institutional Review Boards and Investigators." *Alzheimer Disease and Associated Disorders*, 2004: 171-175.

Appelbaum, Paul S. "Assessment of Patient's Competence to Consent to Treatment." *New England Journal of Medicine* 357 (2007): 1834-1840.

Appelbaum, Paul S, and Thomas Grisso. "Assessing Patients' Capacities to Consent to Treatment." *New England Journal of Medicine* 319, no. 25 (1988): 1635-1638.

Beauchamp, Tom L, and James F Childress. *Principles of Biomedial Ethics.* 6th. New York, New York: Oxford University Press, 2009.

Berg, Jessica W, Paul S Appelbaum, Charles W Lidz, and Lisa S Parker. *Informed Consent: Legal Theory and Clinical Practice.* New York, New York: Oxford University Press, 2001.

Berlin, Isaiah. "Two Concepts of Liberty." In *The Proper Study of Mankind: An Anthology of Essays*, by Isaiah Berlin, 191-242. New York: Farrar, Straus, Giroux, 2000.

Buchanan, Allen E, and Dan W Brock. *Deciding for Others: The Ethics of Surrogate Consent.* New York: Cambridge University Press, 1989.

Cahil, Maura, and Alison Wichman. "Research Involving Persons with Cognitive Impairments: Results of a Survey of Alzheimer Disease Research Centers in the United States." *Alzheimer Disease and Associated Disorders*, 2000: 20-27.

Christman, John. "Liberalism and Individual Positive Freedom." *Ethics* 101, no. 2 (1991): 343-359.

Christman, John. "Saving Positive Freedom." *Political Theory* 33, no. 1 (2005): 79-88.

Crowder, George. *Isaiah Berlin: Liberty and Pluralism.* Cambridge: Polity Press, 2004.

Diekema, Douglas. "Parental Refusals of Medical Treatment: The Harm Principle as a Threshold for State Intervention." *Theoretical Medicine and Bioethics* 25, no. 4 (2004): 243-264.

Diekema, Douglas S. "Parental Refusals of Medical Treatment: The Harm Principle as a Threshold for State Intervention." *Theoretical Medicine* 25, no. 4 (2004): 243-264.

Dworkin, Gerald. "Paternalism." *The Monist* 56, no. 1 (1972): 64-84.

Dworkin, Gerald. *The Theory and Practice of Autonomy.* New York: Cambridge University Press, 1988.

Faden, Ruth R, and Tom L Beauchamp. *A History and Theory of Informed Consent .* New York: Oxford University Press, 1986.

Faden, Ruth R, and Tom L Beauchamp. "The Concept of Informed Consent." In *Bioethics: Principles, Issues, and Cases*, by Lewis Vaughn, 154-158. New York: Oxford University Press, 2012.

Fallon Jr., Richard. "Two Senses of Autonomy." *Stanford Law Review* 46, no. 4 (1994): 875-905.

Fallon, Richard H. "Two Senses of Autonomy." *Stanford Law Review*, 1994: 875-905.

Grisso, Thomas, and Paul S Appelbaum. "Comparison of Standards for Assessing Patients' Capacities to Make Treatment Decisions." *American Journal of Psychiatry* 152, no. 7 (1995): 1033-1037.

Grisso, Thomas, Paul S Appelbaum, Edward P Mulvey, and Kenneth Fletcher. "MacArthur Treatment Competence Study. II." *Law and Human Behavior*, 1995: 127-148.

Guttmacher Institute. "State Policies in Brief: Parental Involvement in Minors' Abortions." 2015.

Hor, Kayhee, and Mark Taylor. "Review: Suicide and Schizophrenia: A Systematic Review of Rates and Risk Factors." *Journal of Psychopharmacology* 24, no. 4 (2010): 81-90.

Kleinig, John. "The Nature of Consent." In *The Ethics of Consent*, by Franklin G Miller and Alan Wertheimer, 3-24. New York: Oxford University Press, 2010.

Lambert, M, et al. "Impact of Present and Past Antipsychotic Side Effects on Attitude Toward Typical Antipsychotic Treatment and Adherence." *European Psychiatry* 19, no. 7 (2004): 415-422.

Leo, Raphael J. "Competency and Capacity to Make Treatment Decisions: A Primer for Primary Care Physicians." *Primary Care Companion: Journal of Clinical Psychiatry* 1, no. 5 (1999): 131-141.

MacCallum, Gerald C. "Negative and Positive Freedom." *The Philosophical Review*, 1967: 312-334.

Maniam, Maya. "Functional Parenting and Dysfunctional Abortion Policy: Reforming Parental Involvment Legislation." *Family Court Review* 50, no. 2 (2012): 241-257.

Meisel, A, L.H. Roth, and C. W. Lidz. "Toward a model of the legal doctrine of informed consent." *The American Journal of Psychiatry* 134, no. 3 (1977): 285-289.

Meltzer, Herbert Y, et al. "Clozapine Treatment for Suicidality in Schizophrenia: International Suicide Prevention Trial (InterSePT)." *Archives of General Psychiatry* (60) 1 (2003): 82-91.

Mill, John Stuart. "On Liberty." In *On Liberty and Other Writings*, by John Stuart Mill, 37-163. New York: Cambridge University Press, 1989.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. "Research Involving Those Institutionalized as Mentally Infirm." 1978.

O.C.G.A 16-5-71 (2012). *Georgia Code.* enacted by Ga. L. 1987.

Palmer, Brian A, Shane Pankratz, and John M Bostwick. "The Lifetime Risk of Suicide in Schizophrenia: A Reexamination." *Archives of General Psychiatry* 62, no. 3 (2005): 247-253.

Section-17A-2. *Alabama Code.* n.d.

Sloble, Alan. "Paternalism, Liberal Theory, and Suicide." *Canadian Journal of Philosophy*, 1982: 335-352.

The American College of Obstetricians and Gynecologists. "Informed Refusal: ACOG Committee Opinion No. 306." 2004, 1465-1469.