Private Rule Following and the Principle of Respect for Autonomy

A thesis presented
to the faculty of
The College of Arts and Sciences of Ohio University

In partial fulfillment
of the requirements for the degree
Master of Arts

Nicholas Smith
May 2015

© 2015 Nicholas Smith. All Rights Reserved.
This thesis titled

Private Rule Following and the Principle of Respect for Autonomy

by

NICHOLAS SMITH

has been approved for

the Department of Philosophy

and the College of Arts and Sciences

Scott S. Carson

Associate Professor of Philosophy

Robert Frank

Dean, College of Arts and Sciences
ABSTRACT

SMITH, NICHOLAS, M.A. May 2015, Philosophy

Private Rule Following and the Principle of Respect for Autonomy

Director of Thesis: Scott Carson

Beauchamp and Childress’s model of autonomous decision making suffers from a Kripkensteinian rule following problem: there must be a difference between meeting the criteria for an autonomous decision, and thinking the conditions for an autonomous decision have been met. Unfortunately, that which determines if the conditions have been met are all purely internal, private facts. Nothing in Beauchamp and Childress’s model as stated calls for anything public. Some public element must be added. The addition of the defense condition from Andrea Westlund’s formal relational theory of autonomy can add this public element without adding in a value-laden definition of autonomy, thereby providing a flexible and practical formulation of Beauchamp and Childress’s model.
ACKNOWLEDGEMENTS

The author would like to thank Scott Carson, Arthur Zucker, and Mark LeBar for their invaluable assistance in crafting this thesis.
TABLE OF CONTENTS

Abstract .......................................................................................................................... 3
Acknowledgements ........................................................................................................ 4
Introduction .................................................................................................................. 6
Three Conditions .......................................................................................................... 8
Problems with the Principle of Respect for Autonomy .................................................. 10
Private Language and Outward Criteria ...................................................................... 14
Relational Autonomy ................................................................................................... 18
The Principle of Respect for Autonomy and Relational Autonomy ......................... 25
Objections ................................................................................................................... 33
The Schematic Solution ............................................................................................... 41
Works Cited .................................................................................................................. 46
INTRODUCTION

In *Principles of Biomedical Ethics*, Beauchamp and Childress propose a principle of “respect for autonomy” as a guiding principle of medical ethics. Beauchamp and Childress rightly note that respect for autonomy runs “deep in the moral community” (Beauchamp, 101). Beauchamp and Childress state three qualities an agent must possess in relation to a specific choice for that choice to be autonomous: 1) intentionality, 2) understanding, and 3) non-control. These three conditions seem\(^1\) to be the necessary and sufficient conditions for autonomy: failing to meet any one of the conditions makes a decision not autonomous, and Beauchamp and Childress “analyze autonomous action in terms of normal choosers” (Beauchamp 104) and note that “This account of autonomy is specifically designed to be coherent with the premise that the everyday choices of generally competent persons are autonomous” (Beauchamp, 104). The criteria that Beauchamp and Childress set out for an autonomous decision align well with intuitions about what makes a decision autonomous. However, as presented, they are entirely useless to the medical practitioner. I will argue that Beauchamp and Childress’s account of autonomous decision making is undermined by a private rule following problem. I will further argue that the beneficial aspects of their account can be salvaged by the introduction of necessarily public features.

First, I will lay out Beauchamp and Childress’s account of autonomous decision making. I will then argue that a private rule following of the sort laid out in Kripke’s *Wittgenstein on Rules and Private Language* is present in Beauchamp and Childress’s account. Finally, I will demonstrate how the most valuable aspects of Beauchamp and Childress’s account can be salvaged by including features drawn from formal accounts of relational autonomy.

\(^1\) I say “seem” here because Beauchamp and Childress do not use the language of necessity and sufficiency.
The project undertaken here is an attempt to resolve (or at least begin to touch on) a real issue in applied ethics. It is concerned with showing how, given a plausible set of conditions for autonomous choice, one can ensure that the criteria are actually met. It is neither an exercise in determining whether one technical definition or another of autonomy actually captures what is meant by “autonomy” in plain speech, nor in arguing for one set of criteria over another. *Principles of Biomedical Ethics* is not an abstract work of normative theory, but a work of applied ethics meant to address the decision making concerns of working medical professionals. The Kripkensteinian private language problem in the definition of autonomous choice renders that definition unusable. No medical professional, no matter how skilled or insightful, can access the private mental content of a patient and determine whether the patient has met the criteria of intentionality, understanding, and non-control. However, the definition Beauchamp and Childress give captures a variety of common, pre-theoretic concerns about autonomous decision making. In general, whether we care for moral reasons (e.g. that autonomous decisions are normatively better than non-autonomous ones) or pragmatic reasons (e.g. one is a doctor hoping to avoid malpractice lawsuits or get a study past a review board), we do generally care that people are informed about serious decisions, that they are not coerced, and that their decision fits with their general life plan. By modifying Beauchamp and Childress’s definition of autonomous choice in such a way as to make the conditions of autonomous choice public, the definition becomes usable for the medical practitioner and policy-maker. It provides real, albeit lean, guidance as to the structure of institutional policies regarding patient choice.
THREE CONDITIONS

Intentionality requires that actions be planned such that there is a “representation of the series of events proposed for the execution of an action” (Beauchamp, 104), and the action must “correspond to the conception of the act in question” though one may not actually achieve the act in question (Beauchamp, 104). Presumably, Beauchamp and Childress state the condition of intentionality in terms of a plan for the achievement of an end and not merely in terms of just an intended end because that is simply too indeterminate. If a cancer patient, when deciding what course of treatment to pursue, only intends “to get better,” it is not apparent that they have intended any particular course of treatment. Anything from faith healing to highly invasive surgery might play a part in achieving such an end, and by intending only “to get better”, the patient has not made clear what they are after. This should be contrasted with a similar case of a patient who intends, say, “to get better by appealing to any means shown in clinical trials to be effective in treating my kind of cancer”. This statement of a plan is vague, but perhaps goes far enough to satisfy the intentionality condition. It includes both a desired end, and (loosely) indicated steps to achieve that end. While still leaving open a variety of treatment avenues, this plan specifies that the patient is open to the use of any treatment within mild constraints, thus indicating the steps towards the end.

Understanding stipulates that “An action is not autonomous if the actor does not understand it” (Beauchamp, 104). One need not have a perfect understanding of the relevant information in order to make an autonomous decision or act autonomously, but merely a “substantial understanding”, as full or perfect understanding would be a prohibitively high threshold. A lack of understanding can be caused by many conditions, such as youth or mental illness, or being given insufficient information by medical professionals (Beauchamp, 104).
Non-control requires that “a person be free of controls exerted either by external sources or by internal states that rob the person of self-directedness.” The kinds of influence that most concern Beauchamp and Childress are coercion, persuasion, and manipulation. Coercion “occurs only if an intended and credible threat displaces a person’s self directed course of action, thereby rendering even intentional and well-informed behavior nonautonomous” (Beauchamp, 138). Persuasion consists in “coming to believe something through the merit of reasons another person advances” (Beauchamp, 139). Beauchamp and Childress state that “approaches that might rationally persuade one patient might overwhelm another whose fear or panic undercuts reason” (Beauchamp, 139). Manipulation is “swaying people to do what the manipulator wants by means other than coercion or persuasion” (Beauchamp, 139). “Manipulative” is vague, though Beauchamp and Childress use informational manipulation “a deliberate act of managing information that alters a person’s understanding of a situation and motivates him or her to do what the agent of influence intends,” (Beauchamp 139) as an example. Coercion, persuasion, and manipulation are all held by Beauchamp and Childress to be at least sometimes acceptable, though “people’s vulnerabilities differ, thereby producing variation in what constitutes an ‘undue’ influence” (Beauchamp, 140). Additionally, “Not all influences exerted on another person are controlling” (Beauchamp, 104) Beauchamp and Childress present these three conditions as the necessary and sufficient conditions a decision or action must meet in order to be autonomous.  

Although they do not use the actual phrase “necessary and sufficient,” it is strongly implied that the conditions are, in fact, necessary and sufficient. For example, Beauchamp and Childress write “We analyze actions in terms of normal choosers who act (1) intentionally, (2) with understanding, and (3) without controlling influences that determine their action.” (Beauchamp 2001, 59) Breaking action down into these three and only these three criteria strongly implies that there is nothing else to consider, which would make the criteria the jseparately necessary and jointly sufficient conditions for autonomous choice.
PROBLEMS WITH THE PRINCIPLE OF RESPECT FOR AUTONOMY

In focusing on autonomous choice instead of the autonomy of agents, Beauchamp and Childress have presented a principle of “respect for autonomy” and described the conditions of autonomous choice without having given a definition of autonomy. This approach contrasts with, for example, Frankfurt’s approach which describes the autonomous agent as one who “is motivated to act as she does because this motivation coheres with (is in harmony with) some mental state that represents her point of view on the action.” (Buss, “Personal Autonomy”), or as one that has “the capacity to evaluate one's motives on the basis of whatever else one believes and desires, and to adjust these motives in response to one's evaluations. It is the capacity to discern what ‘follows from’ one's beliefs and desires, and to act accordingly.” (Buss, “Personal Autonomy”). Christman, for example, argues for an account of agent autonomy on which what matters "in the determination of the autonomy of a desire is the manner in which the desire was formed" (Christman 1991, 10). An agent, on this account, acts autonomously if they act in line with a preference that was formed in such a way that they did not (or would not) have resisted that formation process, their lack of resistance to the process of preference formation was not a result of the influence of factors that inhibited self-reflection, and the self-reflection was at least minimally rational (Christman 1991, 12).

3 Beauchamp and Childress are not the only theorists about autonomy to find that ascribing autonomy to agents as an all-or-nothing property will gloss over the possibility that some decisions are autonomous and some are not. As John Christman notes “Consider the fact that people will, in some aspects of their lives, make decisions autonomously, while in others they are moved by external, heteronomous, factors affecting decision-making. A person with an uncontrollable phobia, for example, may display all of the level-headedness and freedom of thought characteristic of autonomy in aspects of her life not affected by the phobia.” (Christman 1991, 3)

4 I find Christman’s account particularly interesting because it seems to provide a nice bridge from individualist accounts of autonomy to Andrea Westlund’s relational account. Both
Again, Beauchamp and Childress are not focussed on any feature of the agent’s reasoning beyond the scope of a single decision. While Beauchamp and Childress’s account of autonomous decision making certainly resembles some of the aforementioned accounts of autonomous agents, it is not committed to any one of them; it is not committed to a story about the agent, just an account of the decision procedure for a single decision. Indeed, one possible advantage of Beauchamp and Childress’s account is that it could be used in conjunction with a number of different accounts of autonomous agents.

revolve around the process of self-reflection on the desires one acts upon. However, Westlund’s account (at least as I will make use of it) involves the additional step, beyond self-reflection, to defense of one's actions to others, Christman's account involves only reflection on desires and an appropriate process of desire formation. While self-reflection on one's desires is an important part of both accounts, an account such as Christian's does not demand that this self reflection process be evident to others or be anything but internal, and so would not add the necessarily public element we will find desirable.

Additionally, even if it were somehow public (and I suspect a public version of Christman's account would look very much like Westlund's) Christman's account would be less useful as a solution to the problem I raise than Westlund's account for purely practical reasons. Analyzing the decisions of patients on Christman's account may return the result that many of them are acting non-autonomously. It seems highly unlikely that an agent's desires and preferences develop in some way that is completely isolated from the rest of the agent's preferences and desires. Preferences and desires that an agent might use to justify a action might very well be ones that the agent would endorse and whose development they may not have resisted, yet their development might have been influenced by desires and preferences that did not arise autonomously. Indeed, Christman writes that “For a person to be autonomous, it must be true that during the processes where she might have resisted these developing desires, she wasn’t also under the influence of manipulating factors that inhibit the person’s ability to reflect on her desires and those processes that helped form them.” (Christman 1991, 19) These manipulating factors might just be other desires and preferences that were not formed autonomously. It would be incredibly difficult, if not outright impossible to extract those preferences and desires that the agent used to justify their action from the web of the patient's preferences and desires, and show that those preferences and desires did in fact arise autonomously. This process would be, at best, arduous, and more likely inconclusive and likely to leave the agent more confused than when they made their initial choice. It would very likely be not efficient in the least, and therefore not desirable for medical practitioners or patients. Additionally, this process of investigation would itself have to be public, which would only serve to make it even less efficient and possibly more perplexing. By retaining features such as the non-control and planning conditions while working at the level of the particular decision, Beauchamp and Childress's account, when supplemented by Westlund's, is efficient and usable by medical practitioners and patients, while still acknowledging the importance of the patient's actions being in line with their preference.
The three conditions Beauchamp and Childress give for an autonomous decision raise several questions. The lack of definition of autonomy causes difficulties in determining whether somebody actually has met the criteria. For example, regarding the condition of intentionality, it is not clear what constitutes having a “plan of action”. If having a plan of action and making at least a nominal effort to follow the plan is a requirement for autonomous decision making, is there need for a method to ensure that the plan of action comes into being autonomously? Assuming the details of what constitutes a plan are sorted out, the exact role of the plan needs to be made explicit. Does anybody other than the patient need to know about the plan of action, or has the patient met the requirements of the condition if they merely have a plan? With regard to the condition of understanding, it is not clear where the threshold for substantial understanding ought to be set, or how one could justify this threshold. The immediately appealing answer to this question is that one ought to understand enough to be able to act autonomously. Without a definition of autonomy, this is circular. With a definition of autonomy, though, one could appeal to autonomy to provide an indication of the threshold, and a justification for that threshold. Finally, what is the appropriate demarcation between undue controlling influences on the agent, and influences which are acceptable? If Beauchamp and Childress had a definition of autonomy to which they could appeal, it might be clearer which influences would be considered damaging to autonomy.

The above problems present a substantial challenge to the principle of respect for autonomy. Beauchamp and Childress acknowledge that a great deal more specificity is required in order to Resolving them at the level of the principle would add problematic specificity to the principle of respect for autonomy. As Beauchamp and Childress are
well aware of this issue, but are adamant that their theory avoid being overly specific. They note

We acknowledge that our appeal to common morality leaves unsettled problems... Our theory requires that we specify carefully in order to escape abstractness, but we must also not overspecify a principle or rule, which may then become too rigid and insensitive to circumstances. Many specified principles or rules will encounter this problem of too little or too much for some contexts... But without tighter controls on permissible balancing than common-morality theories usually propose, critics will charge that too much room remains for judgments that are unprincipled and yet sanctioned or permitted by the theory. (Beauchamp 2001, 406)

I cannot respond to all of these issues in this paper. An expedient avenue for solving many of the above issues, without adding a lot of detail to the principle itself, would be to appeal to the policy of medical institutions to provide clarification on, for example, exactly what degree of understanding meets the threshold for substantial understanding. However, as a result of a private rule following problem that will be explicated in the next section, institutional policy meant to clarify the above issues will not actually make the three criteria presented by Beauchamp and Childress usable. In what follows, I argue that incorporating necessarily public features of relational theories of autonomy can make it possible for institutional policy to be used to clarify these issues. Beauchamp and Childress’s conditions of intentionality, understanding, and non-control.
PRIVATE LANGUAGE AND OUTWARD CRITERIA

As Wittgenstein wrote, "An inner process stands in need of outward criteria" (Wittgenstein, § 580). We can understand this as meaning that in order for an outside observer to justifiably state that somebody has undergone an inner process or had some inner, private experience, there must be some outward, public sign of this that the observer can check. Drawing on this insight from Wittgenstein, we can see that one of the greatest obstacles in taking Beauchamp and Childress's account of autonomous decision making and turning it into an action guiding tool for professionals is due to the lack of outward criteria for meeting any of the three conditions that Beauchamp and Childress set out. For greater clarity on this, we can look to Kripke's analysis of Wittgenstein's private language argument as found in *Wittgenstein on Rules and Private Language*. The private language argument, as reconstructed by Kripke, concludes in §202 of the Philosophical Investigations with Wittgenstein's statement that "it is not possible to obey a rule 'privately': otherwise, thinking one was obeying a rule would be the same thing as obeying it". Kripke develops this problem through a mathematical example. He considers the case of the addition problem 68+57=125, and asks how we can respond to a "bizarre skeptic" who proposes that "the answer I intended for '68+57' should have been '5'!" (Kripke 8). The skeptic's reasoning is that no fact in the adder's past informs him that the appropriate answer is "125." If the adder has never performed that instance of addition before, then answering "5" to the problem "68+57" would be in perfect alignment with my practice of addition up to this point. Perhaps the rule I was

---

5 I wish to note here that there are numerous interpretations of the private language argument, none of which stands as THE interpretation. I choose to focus on Kripke's here because of a) its relative clarity and b) its focus on rule following.

6 The rule following paradox as stated by Kripke is a very expansive problem. I will not touch on all aspects of it. My focus is not an argument about how words and language have meaning, but purely on the fundamentally public nature of rule following, and the similarities between rule following and criteria meeting.
actually following in the past was something like "all instances of addition involving numbers less than 57 are performed normally, but anything involving a number greater than 57 yields the answer 5." Had I never added numbers over 57 before, no fact in my past would inform me which rule I was really following. With a bit of manipulation, we can see a similar problem for Beauchamp and Childress. The analogous problem, phrased in terms of the conditions of Beauchamp and Childress’s account of autonomy, would look something like this: if all of the conditions for autonomous decision making are purely internal and Beauchamp and Childress’s account of autonomous decision making does not include any necessarily public features, how could one ever be sure that one has met the standards for an autonomous decision? For example, consider the condition of intentionality. The condition of intentionality states that one must form a “representation of the series of events proposed for the execution of an action” (Beauchamp, 104), and the action must “correspond to the conception of the act in question” (Beauchamp, 104). What would it be to form a plan of action privately, resolve to follow it, and say that one had really followed their plan of action? Consider Philosophical Investigations §258:

Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "S" and write this sign in a calendar for every day on which I have the sensation. I will remark first of all that a definition of the sign cannot be formulated.—But still I can give myself a kind of ostensive definition.—How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the sign down, and at the same time I concentrate my attention on the sensation—and so, as it were, point to it inwardly. ... I impress on myself the connexion between the sign and the sensation.—But "I impress it on myself" can only mean: this process brings it about that I remember the connexion right in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'. (Wittgenstein, §258)

We can think of following an action and asserting its adherence to a private plan one has made as "pointing inward" to the plan to show the correctness of the action in the same
way that the diary writer points to an inward sensation in order to assert the correctness of inscribing the “S” in the diary. To point inwardly when giving the criteria for the correctness of action (be it “S” writing or the following of a private plan) is to follow no rule, to meet no criteria at all. It is not only that nobody outside could judge whether one had met criteria, but that purely private criteria are no criteria at all. Kripke’s skeptic might come up and say “Well how do you know you followed the plan? How do you know you had the sensation?” As Wittgenstein says, in both cases, the ultimate plan that one is acting on, the most basic rule one is following, is simply that it “seems right” to the person in question.

What are the outward criteria for the inner processes of meeting the conditions of intentionality, understanding, and non-control? Without including outward criteria in their account of autonomous decision making, Beauchamp and Childress have provided a definition of autonomous decision making that is at best unusable by any medical practitioner, and at worst meaningless: just as there must be a difference between thinking one is following a rule and actually following a rule if rule following is to mean anything, there must be a difference between thinking one has met the criteria for deciding autonomously and actually meeting the criteria for deciding autonomously if the criteria are to effectively ensure respect for autonomy. ⁷

 Particularly prone to this problem are the conditions of understanding and intentionality. In the case of understanding, without some public criteria for meeting the condition, then it is impossible to tell whether one does understand. As Wittgenstein puts it in §201 of the Philosophical Investigations “This was our paradox: no course of action

---

⁷ One might respond that there is a biconditional relationship between condition meeting and autonomous decision making/respecting autonomy: making decisions autonomously is precisely meeting the conditions. It is not that the criteria ensure respect for autonomy as much as they create autonomous decisions. If this is the case, clear and public standards for what counts as meeting the criteria seem, if anything, even more important.
could be determined by a rule, because every course of action can be made out to accord with the rule.” (Wittgenstein §201) Without clear public criteria for the condition of understanding, the criteria are useless, because any assertion of understanding could be made out to accord with the vague condition of understanding. (An analogous case can be constructed for the condition of intentionality.) As I will argue, whatever institutional policy is created to meet the need of this public criteria must also be public in a certain way in order to be sure that the institutional policy reflects the spirit of the criteria for autonomous decision making. That the principle of respect for autonomy does not carry any such public criteria with it is problematic.

A public definition of agent autonomy of the sort provided by Frankfurt, Christman, or Mele could serve as the criteria to be used in determining if somebody had successfully met the criteria for autonomous decision. However, a great advantage of Beauchamp and Childress’s account of autonomous decision making is that it is fairly open regarding the definition of agent autonomy. The criteria of intentionality, understanding, and non-control could be used in conjunction with numerous definitions of agent autonomy. The trick, then, will be to find something public that can be incorporated into Beauchamp and Childress’s criteria for autonomous decision making that is not a full blown conception of agential autonomy. I will propose that features of formal relational conceptions of autonomous decision making can provide a possible solution to problems of this sort.
RELATIONAL AUTONOMY

Relational theories of autonomy, being explicitly public, are particularly well suited to handle problems of privacy. The principle of respect for autonomy that Beauchamp and Childress put forth is plagued by its lack of a public definition of autonomy, which in turn makes the conditions for autonomous decision making difficult to apply. The intentionality and understanding conditions in particular set an entirely internal standard. Relational theories of autonomy, however, rely on human relationships (which are necessarily at least somewhat public) to determine the autonomy of an agent. Incorporating relational aspects into Beauchamp and Childress’s principle of respect for autonomy may be advantageous. The challenge is to find a conception of relational autonomy that does not problematically endorse any particular relationship as the relationship needed for autonomy. This would amount to saddling Beauchamp and Childress with a substantive definition of autonomy, something we ought to avoid in order to retain the spirit of their principle of respect for autonomy.

Relational theories of autonomy are theories of autonomy that, in one way or another, appeal to features of human relationships in order to explain agent or decision autonomy. Such accounts can be divided into roughly two groups, those which are substantive, and those which are formal. Substantive accounts argue that some particular relationships, or features of human life only found in relationships, are those which are necessary for autonomy. One example of this approach is Jennifer Nedelsky’s account in Law’s Relations: A Relational Theory of Self, Autonomy, and Law. She contends that “to be autonomous is to “find one’s own law and live in accordance with it.” (Nedelsky 124). “One’s own law” consists of “commands that one recognizes as one’s own, requirements that constrain one’s life but come from the meaning or purpose of
that life” (Nedelsky 123), where “meaning” and “purpose” are left fairly open.\(^8\) Because people are fundamentally relational, “one’s own law” is not something that is developed in isolation from others. Nedelsky identifies two features of “finding and living in accordance with one’s own law” that make it relational. The first is that “the capacity to find one’s own law can develop only in the context of relations that nurture this capacity” (Nedelsky, 124). If one is an isolated being, one will lack the ability to even find one’s own law, to say nothing of living in accord with it. The second aspect that makes it relational is that the “content” of one’s own law is comprehensible only with reference to shared social norms, values, and concepts\(^9\) (Nedelsky, 124). Ultimately, “We see that relatedness, even dependency, is not... the antithesis of autonomy, but a literal precondition of autonomy and that interdependence is a constant component of autonomy” (Nedelsky, 124-125). Additionally, it is not any particular relationships which make autonomy possible. It is a matter of being in whatever relationships one is in in “the right way.”\(^{10}\)

---

\(^8\) It should be noted here that as much as Nedelsky resists defining autonomy as doing what one wants because it is always possible to question whether what someone wants is something they autonomously want, she has landed herself with the same problem. It is possible to question whether any “core values”, “meaning”, and “purpose” that one has were acquired autonomously. Given Nedelsky’s substantive take on relational autonomy, this seems problematic. As we will see later, Westlund’s formal approach to relational autonomy dissolves some of these issues by requiring only that agents defend their choices with regards to whatever values they possess, with no reference to how they acquired these values.

\(^9\) One possibility: if developing the capacity to find one’s own law requires that one be raised in at least a somewhat communal setting, it is impossible to live in accord with one’s own law as an isolated being, because such behavior would amount to private rule following, which we have seen is problematic. To live in accord with any rule whatsoever, the rule must be public and checkable. Therefore the content of the rule must be the sort of thing that others can understand (hence one’s own law being “comprehensible only with reference to shared social norms, values, and concepts” (Nedelsky, 124)) and one’s own law must be the sort of thing that the public could actually check one on (hence the difficulty in living according to one’s own law as an isolated being).

\(^{10}\) Nedelsky takes the most appropriate model of autonomy to be that of a parent-child relationship because “There we have encapsulated the emergence of autonomy through relationship with others. We see that relatedness, even dependency, is not... the antithesis of autonomy, but a literal precondition of autonomy and that interdependence is a constant
Nedelsky's approach to relational autonomy clearly touches on the Kripkensteinian problem present in Beauchamp and Childress's account of autonomous decision making. For Nedelsky, autonomy is, by nature public. Autonomy does not arise without relationships (which are, of necessity, at least somewhat public), and only arises as a result of relationships. However, this sort of account will not solve Beauchamp and Childress's problem. If one looks at Nedelsky's conception of autonomy, it seems to need conditions of the sort that Beauchamp and Childress have set out in order to explain how any particular decision is autonomous. For example, Nedelsky considers the case of a woman receiving welfare from a government:

A single mother on social assistance is clearly in a relationship of dependence on the state. And despite contemporary rhetoric of empowerment, most such relationships are structured in ways that undermine rather than enhance the autonomy of welfare recipients. Lack of information about eligibility rules, a stance of suspicion from the bureaucracy, invasions of privacy, arbitrary or incomprehensible suspension or reduction of payments keep recipients insecure, off balance, unable to plan ahead, anxious, and uncertain. The payments permit them (barely) to survive, but the way they receive these payments is corrosive of their autonomy. (Nedelsky 39-40)

Clearly, the above described welfare recipient - welfare giver relationship is not conducive to the autonomy of the recipient, and there is certainly an imbalance of power between the recipient and the government. However, Nedelsky’s account lacks a clear component of autonomy” (Nedelsky, 124-125). However, a healthy parent-child relationship is not sufficient for autonomy.

Consider the case of the mother who voluntarily works a demanding job and receives a fair wage in compensation for her labor. One might contend that this woman’s autonomy is reduced in some way as a result of the power imbalance between the employer and the employee. Something about construing the fairly compensated employee as having substantially reduced autonomy seems odd. I take it (if I am interpreting Nedelsky correctly and I believe I am) that the mother on welfare has her autonomy reduced because of her unavoidable dependence on the state (for survival? for a higher quality of life? for something else altogether? Here, Nedelsky is not entirely clear, but we shall assume it is for survival). The implication seems to be a total lack of alternatives to welfare and a non-voluntary entry into the necessarily imbalanced welfare giver-welfare recipient. The voluntary employee at least has the option of that particular job or being on welfare, and (as we have stipulated) is fairly compensated for the work she does. The employer-employee relationship is governed by labor laws, the employee's contract, perhaps
explanation of what constitutes “the right way” to be in any particular relationship. If the right way is just “whatever way is conducive to finding one’s own law” then relational autonomy is defined in a viciously circular way. Clearly, something more explicit is needed. If one tried to describe what it is about the relationship between the woman who is the welfare recipient and the government that impedes the autonomy of the woman, the conditions that Beauchamp and Childress set out would do a very good job. The lack of information regarding eligibility rules, the invasions of privacy and suspensions of payments, and the dependence on welfare payments that prevent meeting the standards of understanding, intentionality, and non-control, respectively.\textsuperscript{12}

While substantive relational theorists such as Nedelsky may contend that Beauchamp and Childress’s conditions could only be met given certain relational preconditions for autonomy, those preconditions will not explain how it is that any particular decision is autonomous. Indeed, it seems that Nedelsky tacitly acknowledges this point: “The very concept of relational autonomy presupposes that autonomy is possible for relational selves; and if that is so, then relationships cannot determine who a person is or what she does or becomes. Otherwise there would be no true autonomy” (Nedelsky 31). On Nedelsky’s account, people are, by nature, relational, but are still individuals. They still must make their own decisions, and the autonomous decision will union regulations, or any number of other social structures that serve to mitigate the imbalance in power between the employer and the employee. (For those interested in expansion of this line of thought see Elizabeth Anderson’s “Expanding the Egalitarian Toolbox: Equality and Bureaucracy” for an extended discussion of relational egalitarian principles in the workplace). The welfare recipient is far more at the mercy of the state than the employee is at the mercy of the employer. Additionally, one might argue that, even if the employer-employee relationship is necessarily imbalanced, gainful employment has the possibility of reducing the effects of undue controlling influences outside of the employer-employee relationship, leaving the employee with a net gain in autonomous decision making capacity in the majority of their life. (This discussion raises the issue of the importance of the type and variety of options in autonomous decision making).\textsuperscript{12} Beauchamp and Childress also have an alternative, fairly obvious rejoinder: accepting welfare payments is required for survival, and being alive is necessary for one to develop any autonomy whatsoever. Therefore, accepting welfare payments is the option more conducive to autonomy than any other option.
(according to our adaptation of Nedelsky’s position) be those made in accordance with one’s own law, but one’s own law can only be discovered or created from the context of whatever relationships one is in.

Relational theories of autonomy, which by definition contain at least some explicitly public features, are particularly well suited to handle problems of privacy. The principle of respect for autonomy that Beauchamp and Childress put forth is plagued by its lack of a public definition of autonomy, which in turn makes the conditions for autonomous decision making difficult to apply. The intentionality and understanding conditions in particular set an entirely internal standard. Relational theories of autonomy, however, rely on human relationships (which are necessarily public, thereby ensuring that all relational conceptions of autonomy carry at least some public element) to determine the autonomy of an agent. Incorporating relational aspects into Beauchamp and Childress’s principle of respect for autonomy may be advantageous. The challenge is to find a conception of relational autonomy that does not problematically endorse any particular relationship as the relationship needed for autonomy. This would amount to saddling Beauchamp and Childress with a substantive definition of autonomy, something we ought to avoid in order to retain the content-neutrality of their principle of respect for autonomy.

An account of relational autonomy that is formal, rather than substantive, may be able to solve the problem of a lack of public criteria and avoid the problems of substantive relational theories, without being completely absorbed by Beauchamp and Childress’s theory. Formal theories of relational autonomy are those that depend on the form or structure of reasoning or decision making as a relational activity. Andrea Westlund proposes a theory in which “autonomy in choice and action… relies (at least in part) on the disposition to hold oneself answerable to external critical perspectives on
one’s action-guiding commitments” (Westlund, 28). It is important to note that Westlund has not set her theory up as a complete theory of autonomy. The disposition to hold oneself answerable to external critique is not the necessary and sufficient condition for autonomy. It is merely the necessary condition (Westlund, 28). This means that it is possible to fill out the remainder of the necessary and sufficient conditions with pieces of another theory of autonomy.

I contend that we can use a slightly altered version of this condition in order to add a public feature to Beauchamp and Childress’s principle of respect for autonomy that does not bring with it a full-blown definition of autonomy.

Westlund’s formal relational theory of autonomy is useful for several reasons. It avoids endorsing any particular relationship or set of relationships as the set necessary for being autonomous. Additionally, her theory can handle a variety notoriously tough cases for many conceptions of autonomy, such as those of women who seem to have chosen life in an “oppressive” religious community, or those of the “happy slave”. A substantive theory of relational autonomy (i.e. one that endorses a particular relationship as the source of autonomy) might declare that members of such a community lack autonomy (or lacking all but the lowest degrees of autonomy), even if they insist that they genuinely chose their life. There seems to be something intuitively wrong about denying autonomy to those who offer robust defenses of choices, even if their choices are in line with values we might find “oppressive”. Westlund’s theory allows for the possibility that members of a supposedly oppressive community could offer a suitably strong account of their religious choices, strong enough that an external critic could recognize the community members as having made an autonomous choice, even if the choice rests on values the critic does not hold. Westlund’s theory is also value-neutral: as long as one is disposed to defend one’s choices and the values they depend on to
external critics, one can defend any choice or value. One does not need to find a way to
determine that some particular relationship or relationships have met some particular
degree of “quality” and there is no need to appeal to circular metrics of relationship
health.

Here, one might contend that Westlund has only pushed the problem back a
level: now that people can be considered autonomous as long as they offer or are
disposed to offer a defense of their choices on the grounds of whatever values they hold,
do we have any grounds to critique particular values, or is it the case that “anything
goes”? It is beyond the scope of this paper to give a full account of how we either justify
our values to others, or of how we find grounds to justifiably critique the values of others.
To offer a short, and by no means exhaustive, suggestion, it seems to be the case that
the values people will accept as good grounds will depend an awful lot on the cultural or
social framework one is in. Piety and faith in a loving and benevolent deity will likely be
seen as good values in heavily Christian society, for example. In order to maintain the
value neutrality of Westlund’s account, I propose that the focus of her formulation of
autonomy be on defending one’s action rather than the particular values one defends
them with. Obviously, this does not solve the problem entirely, but it seems to go well
with the use of Westlund’s account to solve tough cases such as those of women in
nominally oppressive religious communities: obviously, most of us do not share the
values those women do, but we can look at their defense of their decision on the ground
of those values. In considering the case of somebody who autonomously decides to do
or go along with something terrible (perhaps think of a German soldier in WWII who
autonomously chooses to do the best he can at exterminating Jews), we need not
endorse the choices that the person makes (to kill Jews) to recognize that this person is
autonomous.
THE PRINCIPLE OF RESPECT FOR AUTONOMY AND RELATIONAL AUTONOMY

A formal relational conception of autonomy has a chance at being both attentive to relational concerns, and structured in such a way that it can mesh well with an individualist theory like Beauchamp and Childress’s. By requiring that any decision maker have the disposition to defend their decision to others, Westlund’s theory acknowledges that humans are relational beings, while still ensuring that the focus is on the individual decision maker. It seems that Westlund’s theory could bring a genuinely relational aspect to Beauchamp and Childress’s individualist theory of autonomy.

Additionally, unlike substantive theories, Westlund’s theory need not cease to be constitutively relational when it is fused with Beauchamp and Childress’s. As shown above, if one were to combine Nedelsky’s account of relational autonomy with Beauchamp and Childress’s account of autonomy, Nedelsky’s is completely absorbed. Including that sort of account does not show in any way what about any particular decision is autonomous or non-autonomous, nor does it add anything public to Beauchamp and Childress’s account. Westlund’s defense condition, however, does not get absorbed by Beauchamps and Childress's account. Westlund contends that "Autonomy, on [my] view, requires an irreducibly dialogical form of reflectiveness and responsiveness to others. But this type of relationality, while constitutive, is formal rather than substantive in nature and carries with it no specific value commitments." (Westlund, 28) The defense condition does not pick out features of a decision that are already in Beauchamp and Childress’s criteria, while it does add a necessarily public element.

What, then, is the result when one does attempt to bring Beauchamp and Childress’s theory and Westlund’s theory together? To determine this, we should consider the conditions each has set for autonomous decision making and examine the resulting interactions. Beauchamp and Childress set intentionality, understanding, and
non-control as the necessary and sufficient conditions of autonomous decision making. Westlund set a disposition to defend one’s choices and values to external criticism (call this the defense condition) as a necessary condition of autonomous decision making. If Westlund’s theory does need some filling out, then Beauchamp and Childress’s conditions of autonomous decision making may fit the bill if one is willing to give up on the three conditions as the sufficient conditions. Overall, this seems a small price to pay in order to salvage the three conditions. 

In order to make use of Westlund’s theory of relational autonomy, we will need to modify it slightly yet still retain the constitutively relational nature of her theory. Westlund problematically states her theory in terms of a disposition. If the theory is not somewhat modified, it will be not provide an escape from the "Kripkensteinian" privacy problem. As Westlund notes,

> I do not think that one must look outside the agent’s psychology... to overcome the problematic kind of internality. Instead, we must consider whether the agent is limited, in her reflective capacities, to essentially monological... This distinction does not map onto any straightforward distinction between what is internal and what is external to the agent. The disposition to hold oneself answerable to others is, after all, a feature of the agent’s psychology, and thus internal to the agent. But it is thus nonetheless a disposition to be engaged by what is external to the agent... (Westlund 33)

Westlund’s theory has exactly the same privacy problems as the conditions set by Beauchamp and Childress. Westlund is concerned with giving an account of what it is for a decision to be autonomous rather than of how outsiders can judge a decision to be autonomous, and so is perfectly content to appeal to private features, such as dispositions, in developing her account of autonomous decision making. However, adding

---

13 In examining the interactions between the defense condition and Beauchamp and Childress’s theory, I will not deal with the condition of non-control because the bulk of the controlling influences Beauchamp and Childress are concerned with are external influences, and are therefore necessarily public.
more privately satisfied conditions to Beauchamp and Childress's list of privately satisfied conditions will not make any of them public. No matter how much internal dialogic reflection a patient goes through, none of this can be accessed by any outside observer, short of the patient explaining and recounting their reflective process and opening it to scrutiny. The most promising solution (short of foisting a necessarily public definition of autonomy onto Beauchamp and Childress or Westlund, all of whom have presented a value neutral conception of autonomous decision making) is to make a small modification to Westlund's theory to bring relational elements, the irreducibly dialogical form of reflection" (Westlund 28) into the public sphere. The only way to do this is to state Westlund's theory without referring to private features of people, such as dispositions or internal reflection.

I believe it is reasonable to recast Westlund's theory of autonomy in terms of actually defending the choices one makes. When an observer is attempting to consider whether some decider's particular decision is autonomous on Westlund's account, it is reasonable for the observer to engage the decider in discussion about their reflective process, to criticize their view, or to ask them to defend it. In giving examples of the application of her theory, Westlund herself does this. Describing how a Taliban woman might be autonomous in her choice to remain in what could be considered an oppressive community, Westlund writes that "to treat her as non-autonomous even as she speaks on behalf of her self subordinating commitments…" (Westlund 29). This indicates that an actual defense of one's choices plays an important role in the disposition to defend oneself to critics. Additionally, if we think of a disposition as a general tendency to do X given conditions Y, it would seem awfully strange to say that somebody had a disposition to defend themselves to their critics if, when faced with criticism, they never defended their choices. Henceforth, Westlund’s defense condition will be used in this
modified form, that of actually defending oneself to external critics. This condition is, of course, not perfect. What counts as a “sufficient defense” will be a question left perhaps to community values and standards, or institutional policy. The one stipulation that seems reasonable to make is that whatever the standards are, they must be as close to value-neutral as is possible, in order to maintain the value-neutrality of Westlund’s position.\textsuperscript{14}

The condition of understanding fits fairly easily with Westlund’s condition of defense. It is difficult to believe that anybody could offer a suitable defense of a decision they have made without at least some understanding of what they have chosen, as well as a good understanding of the values behind their decision. Obviously, a perfect understanding of the relevant information is unlikely to ever actually be achieved, but the standard of “substantial understanding” is, prima facie, reasonable.\textsuperscript{15} The defense condition could be used to set the threshold for “suitable understanding”. One has a suitable understanding of the relevant information when one is able to appropriately integrate the information into a defense of their actions. One cannot reasonably expect the average cancer patient to learn the minutiae of oncology, chemistry, insurance law, and hospital policy prior to consenting to a course of treatment. However, one might expect that a patient would learn at least the details of the available treatment options, the possible payment plans for each treatment, and the methods the care providing institution uses to carry out the various procedures. These are the sorts of details one

\textsuperscript{14} It strikes me as likely that the criteria for any sort of defense condition will be directed towards ascertaining the coherence of the agent’s choice with the agent’s values. One may object to this on the grounds that we will need to know what autonomy is in order to give suitable criteria for public defense. However, it is possible to read Beauchamp and Childress (and, by extension, my modification of Westlund) as essentially saying “Any characterization of autonomy (the property of an agent) will do. Simply plug that definition of autonomy (the property) into the conditions for autonomous choice and set the standards for public defense accordingly”.

\textsuperscript{15} Admittedly, “substantial understanding” is a vague standard, but the exact criteria can perhaps be reasonably be pushed to the institutional or legislative realms.
would need in order to be able to defend a treatment choice. By integrating such details into a defense of a treatment choice, the patient would demonstrate that they had a substantial understanding of the relevant information.

Obviously, the defense condition cannot not solve all of the issues surrounding the condition of understanding. While the defense condition might be used to guide policymaking in setting the minimal threshold for understanding, it would not, for example, necessarily ensure that someone did not have a fundamental misunderstanding of the information. It is possible (though it seems unlikely) that a patient could fundamentally misconstrue relevant information but manage to convey their misconstrual in such a way that they appear to adequately understand the information. In this case, the depth of the defense and the variety of the information integrated can help indicate that a patient has a substantial understanding, though ultimately, it could never be guaranteed to indicate with perfect certainty that the patient’s choice was made with full understanding. Obviously, meeting the defense condition will require that there is some degree of reasonably in depth communication between the patient and some hospital representative (after all, signing a consent form hardly requires defending one’s choices). However, I take it that meeting Beauchamp and Childress’s conditions would require some level of communication anyway. What the defense condition changes is the degree of communication, not that there is communication. Additionally, a suitable

---

16 One potential problem with this formulation is the issue of deferred decision making. It is not clear how much understanding (or degree of non-understanding) of a scenario one would need to have in order to decide that they do not want to make the choice themselves/ought not make the choice themselves. Presumably, the standard would be that one needs to have enough understanding of a situation to know that one would rather have somebody else deal with it. That said, the problem of how to ensure that autonomous decision making can be deferred, especially in the case of those who do not wish to know about their condition, is not unique to Beauchamp and Childress’s or Westlund’s accounts of autonomous decision making: any account of autonomous agency or autonomous decision making must tackle this issue, and it is not clear that any “winner” has been determined in this debate.
defense of a choice would not ensure that somebody actually had all of the relevant information. Were a physician to fail to tell their patient about a serious side effect of their treatment, the patient could not integrate this information into their defense of their choice. The defense condition will merely set a threshold for, and serve as an indicator of substantial understanding, assuming that a decider does have a correct understanding, and if they have all of the relevant information.

The condition of intentionality could also be improved by incorporating the defense condition. One of the most obvious problems with the defense condition is that it is stated in terms of “having a plan”. This does not entail that the plan is in any way public. As it is stated, the intentionality condition could be met privately, by merely having a plan. This, however, leads directly to the private rule following problem: there must be a difference between thinking one has followed a rule and actually following a rule, and a rule cannot be followed privately. If meeting the intentionality condition consists in having a plan with a series of events designed to achieve some end, then in order for there to be a difference between thinking one has met the intentionality condition and actually meeting the intentionality condition, there must be a difference between thinking one has a plan and actually having a plan. The difference between thinking one has a plan and actually having a plan is likely to be one of depth or detail. Beauchamp and Childress define a plan as a “representation of the series of events proposed for the execution of an action” (Beauchamp, 104). One could imagine being absolutely sure they had a series of events in mind that would allow them to achieve an end, only to find out, upon making the plan public, that they had missed crucial steps, misunderstood how to move from one step to another, or otherwise failed to actually make their plan “correspond to the conception of the act in question”. This would tie the intentionality condition (at least mildly) to the understanding condition.
Additionally, if we understand meeting the intentionality criteria to be like rule following, and actually having a plan is required to meet the intentionality criteria, then plan-having must have at least some public feature. Consider a cancer patient who has a plan to use all and any treatments to improve their health. If this plan is never expressed and consists in a purely internal formation of a series of actions designed to achieve an end, then it is entirely unclear whether the patient has a plan. Nobody could know that the patient endorses using any and all treatments available to improve their health. In fact, it could never be known whether the patient has a plan at all. An assertion by the patient that they have a plan, without an expression of the plan itself, will do nothing to aid in meeting the defense condition. If our patient were to say “Yes, I have a plan to treat my cancer” without a public expression of the plan itself, it could not be known solely from that statement whether the patient has actually formed a plan, or whether they merely think they have formed a plan. The only way to ensure that the patient has actually formed a plan, then, is for the plan itself to be public. This is where the defense condition comes in. If one states and defends their decided end and the series of actions that they believe will lead to that end, then it becomes possible judge whether the patient actually has a plan. In the case of our cancer patient, their merely having a plan to use any and all treatments shown to be effective is never sufficient for knowing that they have a plan, but the patient’s statement “Yes, I have a plan to treat my cancer: I endorse the use of any and all treatments shown in clinical trial to be effective at treating my sort of cancer because I wish to be well at any cost” would indicate that the patient actually had a plan. Here, as with the condition of understanding, the defense condition will not actually provide a criterion for what will count as a plan. That question is one that perhaps can be pushed back to the level of institutional policy or legislation. It could also be solved by an appeal to a public definition of autonomy, but no such
definition is forthcoming from Beauchamp and Childress. Additionally, were the agent to leave some critical aspect of their plan out of the defense, or lie about their plan, then even with public criteria of some sort for a plan, one could not know if the patient had a plan. Perhaps we might wish to stipulate that any suitable defense of a plan will be one that is both full and honest. This, of course, faces its own privacy problem that may be in principle insurmountable. Still, in a suitably abstract case, the stipulation would do. Additionally, the honesty of the participants will be an issue no matter how one defines autonomous choice or writes institutional policy. Without committing Beauchamp and Childress to a particular definition of autonomy, then, the publicity that the defense condition forces offers about as much progress as is possible. Given any particular definition of autonomy or institutional/legislative rule about what counts as a plan, then the defense of an agent’s plan by that agent will ensure that the plan is public enough to be judged against those criteria.
OBJECTIONS

If joining the defense criterion to each of Beauchamp and Childress’s criteria merely places the demand for specifics on institutional policy, why bother with the defense condition at all? If following institutional procedure is already a necessarily public practice, what role does the defense condition play? The purpose of including the defense condition is to place some minimal restrictions on the structure and application of institutional policy in order to ensure that institutional policy reflects the conditions of autonomous decisions. The defense condition does this by ensuring a certain degree of engagement between patients and practitioners. To understand how the inclusion of the defense condition would restrict institutional policy, we can look at the clearest example, the condition of understanding. Most medical institutions (certainly any reputable one) would have some institutional policy meant to nominally ensure that patients in unexceptional circumstances have some degree of suitable understanding of whatever procedure they were deciding to undergo. (By “unexceptional circumstances”, I mean to exclude deciders who, for example, have mental conditions such as severe dementia which would preclude having a particularly deep understanding of most circumstances, who have appointed a proxy, are comatose or otherwise unable to communicate, or are not undergoing emergency treatment that would make a defense time-prohibitive.) Imagine, then, that the institution’s policy for ensuring that a condition of understanding has been met amounts to nothing more than a cursory question by a doctor or administrator (“Do you understand this procedure and its risks?”) or the mere

---

17 Given that Beauchamp and Childress’s conditions of autonomous decision making are taken from Principles of Biomedical Ethics, I have mostly used the language of patient-physician relationships and examples pulled from medicine when discussing Beauchamp and Childress’s theory. However, it seems plausible that my general argument applies to decision making in a variety of contexts, or at least those contexts in which Beauchamp and Childress’s conditions are applicable.
signing of a form (“Sign here to indicate that you understand the procedure you and that you consent to it”). Such a practice does nothing to ensure that the condition of understanding is actually met. If we accept, as the Kripkensteinian problem demands we do, that it is perfectly possible for one to seem as though they understand and that there must be a difference between actually meeting the condition of understanding and seeming to do so, then it is unclear how the answer to a mere yes-or no question or signing of a form would be suitable for ensuring that the patient actually meets a condition of understanding. When asked “do you understand?” the person who thinks they understand (but does not actually understand) will answer in the affirmative. Clearly, such a question is insufficient. By demanding a more in depth and active defense of the patient’s professed understanding, it becomes possible to judge whether the patient actually does have some understanding. The more communication between patients/deciders and physicians, the more sure one can be that the patient has actually met whatever criteria for understanding has actually been set.

The defense condition places demands on the structure of institutional policy, but not necessarily the content. It does not set a threshold for understanding. Instead, it requires that the procedure for ensuring that the threshold is met is public and can therefore ensure that the threshold is actually, and not just seemingly met. One possible issue is that an institution will have a policy which sets the threshold for understanding so low that a yes-or-no question would suffice for an indication of understanding. Such a policy would seem to violate the spirit of the understanding condition and make the understanding condition entirely trivial. In this case, the defense condition will be of no

18 This is true regardless of the standard that has been set. Even if a very low threshold for understanding has been set (e.g. the patient only needs to know the name of the procedure, projected recovery time, and side effects), more communication between the patient and the physician can only serve to aid in demonstrating that the patient has actually met the condition.
use whatsoever. Rectification of this sort of problem would likely have to come from some sort of review of the institution’s definition of “understanding”. With a more detailed definition of understanding, the defense condition would be better able to inform the structure of policy.

In reformulating Westlund’s condition as requiring an actual defense of one’s choices and not merely a disposition to defend ones choices, I have opened myself to a fairly sizable objection. In requiring that actors actually defend their actions in order to be judged to have acted autonomously, it seems that we only act autonomously when we defend our actions. This clashes somewhat with our everyday notions of autonomous action. When I wake up in the morning and decide to make coffee and bacon instead of tea and toast, I don't speak to anybody else, or even see anybody else. Nobody, however, proposes that my choices and actions with regard to my choice of breakfast were non-autonomous. If left as a disposition, the defense condition would not cause this problem because it seems perfectly reasonable that, if I were questioned about my breakfast, I could make some sort of defense of my choice and action. When formulated as requiring an actual defense and not merely a disposition to defend, this option is not available.

At some level, this problem simply must be accepted. If my solution to the private rule following problem in Beauchamp and Childress’s account of autonomy is to be useful to medical practitioners, then an actual defense is likely to be the most efficient and practical way to add publicity. However, the problem may not be devastating. First, when a medical professional professional is concerned with determining whether a patient has made an autonomous decision, the professional is concerned not just with whether the patient has made an autonomous decision, but with whether he can make a reliable judgement that the patient has done so. If the professional uses only the
conditions as stated in Principles of Biomedical Ethics, they will not be able to make such a judgement. The addition of the defense condition makes it possible for the professional to make a more reliable judgement by compelling the patient to make public at least some features of their private mental content and reasoning that went into meeting the other three conditions.

Additionally, some of the above concerns can be sidestepped by the creation of social or institutional rules and conventions under which we can assume, or agree to act, as though certain actions are autonomous. As Kripke notes

The criterion by which others judge whether a person is obeying a rule in a given instance cannot simply be his sincere inclination to say he is... However, after the community judges (based on the original criteria) that he has mastered the appropriate rule, the community may (for certain rules) take the subjects sincere claim to follow it in this instance as in itself a new criterion for the correctness of this claim, without applying the original criteria.” (Kripke 101-102, footnote 82)

We need not, for every decision we make, go about defending ourselves to other people or risk facing charges of having acted non-autonomously. It is reasonable to accept that for a wide variety of decisions, people can simply be taken to be acting autonomously, or taken at their word if they assert they are. There is no need for me to defend my decisions about my breakfast. Under normal circumstances (e.g. I am not imprisoned or captive, I have the mental faculties of a normal adult, I am not intoxicated or otherwise impaired, etc) they can be treated as though they are autonomous without an explicit defense simply because a) I am judged to be an adult and as such am generally considered to be able to make such decisions and b) there are generally no serious repurcussions following from my decision to make bacon instead of toast. My contention is that there are, however, good pragmatic reasons for not taking as autonomous many, although probably not all, medical decisions. Practitioners and medical institutions have an interest in ensuring that patients have made their decisions
autonomously. If patients are under the impression that they have been deceived, ill
informed, or coerced into certain decisions, especially those with weighty consequences
(e.g. invasive surgery, courses of medication that have serious side effects, procedures
with a long inpatient phase or weighty financial burden), then there may be grounds for
legal action against the practitioners or institutions. This may be the case even if the
physician or institution is convinced that they are acting in the patient's best interest, or
even if the patient is ultimately convinced that the course of action they were coerced
into was in their best interest.19 As such, physicians and medical institutions should be
open to practices that enhance the reliability of autonomy judgements about patient
decision making. As Kripke's footnote suggests, it is likely not necessary to elicit a
lengthy defense from every patient for every decision. If a patient, for example, assented
to being given a bandage for a minor wound, or having their blood pressure and
temperature checked, it seems unnecessary to require that the patient go in to great
detail about how their assent fits in with some larger set of values and life goals. The
potential legal repercussions for the hospital and practitioner are so small and unlikely
that it would simply not be worth their time. Where to draw, the line would likely be a
matter for hospital administrators, ethics boards, and medical associations to decide.20

There are two additional objections that must be considered. First, it is possible
that Westlund's defense condition will make it possible for one to defend a decision with
respect to values that were not autonomously acquired. This is the sort of case that

19 This is separate from the issue of the physician acknowledging that a patient has made an
autonomous choice and then choosing to ignore it.
20 A similar rejoinder (leave it to the institutions and ethics boards) is, I believe, a reasonable
response to the question of “what counts as a defense”. Again, the purpose of this essay is
not to write policy or argue for one particular account of agent or community/institutional
values as being “correct” but merely to show, regardless of the values an agent or
community/institution holds, the features that institutional policy must include in order to
ensure that policies and rules governing autonomous decisions on the part of patients are
practical and usable.
Christman's historical account of personal autonomy is meant to rule out as autonomous. One might consider the (admittedly uncommon) case of some person, call them P, raised in a cult that brainwashed its members into, among other things, refusing liver transplants. P leaves the cult, but cannot quite shake either the feelings of guilt that come with breaking the rules of the cult or some of their basic beliefs, even though he knows he no longer bound by them. P develops cirrhosis later in life, and will need a liver transplant to survive. P would like to live, and acknowledges that the transplant is vital to his survival. However, the feelings of guilt are too strong and P cannot bring himself to assent to the procedure, defending his choice on the grounds of the odd beliefs he holds as a result of his brainwashing, despite no longer being a member of the cult.

A few things need to be said about the above case and similar cases. First, if the patient's beliefs were acquired non-autonomously, it is still possible for the patient to endorse them. Perhaps the patient views them as adding some larger meaning to their life, or finds that they add structure to an otherwise disorganized existence. In this case, the patient's decision ought to stand.

Second, if the patient does not endorse the beliefs and principles they use to defend their action, this could be viewed as a sort of temporally extended manipulation. Being unable to avoid adherence to religious tenets that one technically no longer endorses (for whatever reason, be it brainwashing, lingering guilt, or something else) is exactly the sort of circumstance that the non-control condition is meant to cover. A decision made under such constraints would violate the non-control condition, and would not be considered autonomous, regardless of the value neutrality of the defense condition. If the patient acknowledges that they feel manipulated by their non-autonomously acquired beliefs, then there are ways that they might rectify the situation. If the problem is merely that they themselves are unable to assent to a procedure but
acknowledge that it would be in their best interest, then perhaps granting a trusted third party power of attorney would allow the patient to get the outcome they desire without the feeling that they are personally defying some principle they hold, or without feelings of guilt. Alternatively, a guardian appointed by a third party might decide for the patient using a standard such as the "patient's best interest." Finally, physicians have pragmatic restraints to consider. At a certain point, a decision to go ahead with or abandon a certain procedure will simply have to be made. In some cases, it will simply be impossible to delve deeply into the patient's personal history in order to determine whether the beliefs they use to defend their decisions were formed in such-and-such a way. As shown in the response to first objection in this section, if the patient has met whatever institutional and communal standards are set for being an autonomous decider, there may come a point at which the appropriate action is simply to take the patient at their word that the beliefs they use to defend their decision really are their belief.

The second objection is that the extension of Beauchamp and Childress's conditions of autonomous choice to include a defense condition is unnecessary because clearly, any actual application of the principle of respect for autonomy is going to require some sort of interaction between the patient and the medical institution in some way, be it through their doctor, an institutional administrator, or somebody else. Making a decision in total isolation just wouldn't happen. While this is true, I do not believe it represents a substantial objection to my position. The demand of the defense condition is not merely

---

21 Best-interest standards are rife with problems, most of which are variations on the question "is there an objective standard by which the patient's best interests are be judged?" If the answer to this question is "no" or the answer relies on something private, then such standards are of little use. Given that Westlund's account of personal autonomy points towards (although perhaps does not outright endorse) a subjective account of human good, it is not clear how useful a "patient's best interest" standard would be.
for interpersonal interaction, but for interaction of a certain sort: that in which the patient defends their choice of action to outside parties. The defense condition places some minimal restrictions on what sort of interactions between patients/deciders and physicians/institutions/decidees are suitable for ensuring that the intentionality, understanding, and non-control conditions are met. Interactions in which the outside parties do not take any time to engage the patient and simply ask them to, say, sign a form indicating asent to a procedure, or merely record the patient's choice of action, are not likely to constitute a defense of ones choice.22 At a bare minimum, some degree of discussion about why the patient made the choice they did, or how they align their choice with their belief would seem to be required. Additionally, to contend that the defense condition is a meaningless extension of Beauchamp and Childress's three principles is to simply ignore the private rule following problem. The point is that not just that one could not tell if a patient had met Beauchamp and Childress's three conditions if there were no outward criteria. It is that without public criteria and checkability, there is nothing that actually is meeting the conditions. As Wittgenstein says, "it is not possible to obey a rule "privately"; otherwise, thinking one was obeying a rule would be the same thing as obeying it." (Wittgenstein §202) For a practice to really be criteria-meeting, it must be subject to some sort of public check. Meeting criteria privately is not really meeting criteria because there is no assurance that what one has actually done meets the criteria.

---

22 This will not be a problem for all cases. For example, if the patient and practitioner are on the same page, so to speak, then there will be nothing to defend.
THE SCHEMATIC SOLUTION

Beauchamp and Childress’s definition of autonomous decision making incorporates three conditions: understanding, intentionality, and non-interference. I have argued that definition of respect for autonomy/autonomous decision making is subject to a rule following paradox. One can escape from this paradox by ensuring that the conditions of autonomous decision making are met publicly. I propose that one way of doing this would be to add a fourth condition, a condition of defense. The defense condition ensures that the criteria for autonomous decision are met by forcing the decider to meet the conditions of intentionality, understanding, and non-control publicly. While the defense condition can not solve all potential problems with Beauchamp and Childress’s definition of autonomous decision making, I believe it provides a substantial improvement.

Kripke’s solution to the private language problem is essentially that community checkability of the results of purported rule following allow us to determine if others are, indeed, following rules:

The solution turns on the idea that each person who claims to be following a rule can be checked by others. Others in the community can check whether the putative rule follower is or is not giving particular responses that they endorse, that agree with their own. The way they check this is, in general, a primitive part of the language game… (Kripke, 101)

My solution to private criteria issues in respect for autonomy is analogous to Kripke’s solution to the private language problem in that it “depends on agreement, and on checkability-on one person’s ability to test whether another uses a term as he does” (Kripke, 99). However, while Kripke’s solution looks to public checkability and agreement on the use of terms to solve the skeptical problem, that solution is not directly open to us. “Understanding”, “non-control”, and “intentionality” are loosely defined, leaving far too
much room for disagreement and interpretation for merely looking towards agreement to
be a practical solution. Moving one level up and looking towards public agreement for
use of the term “autonomy” as a way to flesh out “understanding”, “intentionality”, and
“non-control” also seems likely to fail. It will likely result in either forcing a definition of
autonomy on Beauchamp and Childress (which we are explicitly trying to avoid), or lead
one towards circularity. By amending the conditions set by Beauchamp and Childress to
include the defense condition and measuring the public defenses of choosers against
public, institutionally defined metrics, the understanding, non-control, and intentionality
conditions of autonomous choice become practical and usable guidelines for institutional
policies regarding decision making and real criteria for judgments regarding everyday
decision making, while still being flexible enough to accommodate (at the level of
institutional policy) a variety of concepts of autonomy. Instead of going by Kripke’s
solution and looking for community agreement regarding the appropriate use of terms,
one will look for agreement by the appropriate judges23 (e.g. physicians, nurses, hospital
administrators)24 about whether a decider has met institutional criteria.

The defense condition, of course, no matter how one integrates it into
Beauchamp and Childress’s conditions, does not actually provide a full-blown remedy to
the issues in Beauchamp and Childress’s conditions for autonomous choice. Certainly, it
is clear that adding public features to the definition of autonomous decision making is
necessary to make the definition a usable one. The defense condition (when stated in
terms of actual defense and not a disposition) does do that. However, the defense

23 It is important to keep in mind the following: “The criteria by which others judge whether
a person is obeying a rule in a given instance cannot simply be his sincere inclination to say that
he is; otherwise there would be no distinction between his thinking he is obeying a rule and his
really obeying it (PI paragraph 202)

24 The preceding list is meant to keep with the medical theme of this paper. This could be
generalized to a variety of professions.
condition alone is not *sufficient* to make Beauchamp and Childress's definition of autonomous choice a fully functional, practical definition that medical practitioners can use to guide choices.

What the defense condition provides, instead of a detailed solution or metric that medical practitioners could universally apply to particular situations in order to determine whether or not a decision is autonomous, is provide a solution schema. What requiring an actual defense does is provide a minimal foundation on which to begin structuring the remaining requirements of autonomous decision making. Each condition of autonomous choice that Beauchamp and Childress provide brings up a variety of issues that will be relevant when judging whether a patient’s decision is autonomous. For example, a definition of understanding will be needed in order to formulate clearly what will count as “substantial understanding” and the nature of a defense that would signify significant understanding. Definitions of “plan” and “representation” will be needed in order to determine what it is to have a representation of events leading to a desired end, and how to determine what would be needed in a defense to show that one has a plan. Even more basic will be the need for an explication of what counts as a defense.

While this essay has focussed merely on the definition of autonomy Beauchamp and Childress provide, it is important to remember that the definition is only one part of a larger principle of respect for autonomy. Therefore, once all the other problems in each of the different conditions has been cleared up, if a principle of *respect for autonomy* is to be a useful principle in biomedical ethics, there needs to be a clear, public definition of what it is to *respect* some person’s autonomy. All of the above issues, and related ones, will require more basic answers to tough questions in epistemology (especially regarding the nature of justification) and in metaethics (regarding the nature of moral reasoning and its relationship to epistemic reasoning). Given the highly contentious nature of the
problems that need to be solved in order to both develop a clear definition of autonomous decision making and to incorporate it into a principle of respect for autonomy, it seems highly likely that any set of clarifications and definitions provided that was general enough to be widely accepted would be too vague to provide action guidance, and any set of clarifications that was specific enough to provide action guidance would most likely be highly contentious. The most feasible way to solve these issues is likely to push the definitions back to the level of institutional policy, community standards, or both.\textsuperscript{25} What might it look like to apply a sort of schematic solution to a particular case of a decision made by a patient? Although this solution is not yet fully developed, I believe it would look something like the following. In any particular case of patient decision making, there are two general issues at stake: is the patient’s decision autonomous, and how should the institution ensure that they respect the autonomy of the patient. For now, we will apply the schematic solution to the first question.

The schematic solution will include reference to the conditions of intentionality, understanding, non-control, and require that certain “blanks” be filled in” by different communities and institutions in a way that is in line with the defense condition. The intentionality condition requires that the patient have a series of events or processes in mind for the achievement of a certain end state, so the institution will need to ensure that their policy a) provides information about the degree of detail the patient must have in their plan in order to count as a plan, and b) mandates that the patient express the plan above the specified threshold of detail and demonstrate, if pressed, how this plan fits with their desired end. The understanding condition requires that patients possess a substantial understanding of their choice, so the schematic solution demands that

\textsuperscript{25} One might look at Robert Brandon’s “Adaptation and Evolutionary Theory” as an example of how one might pursue a schematic solution to certain definitional problems, though his of course deals with adaptation and fitness, and mine with questions about autonomy.
institutional policy at least a) define “understanding,” b) set a threshold for how much understanding must be demonstrated by the patient for their understanding to be considered “substantial,” c) define which information will be considered relevant to which sorts of decisions, and d) set forth a procedure by which the patient actually demonstrates their command of relevant information. Perhaps this will be woven into the intentionality condition and the expression of the intended plan, or it may come in the form of a survey or conversation with physicians or hospital administrators. The defense condition will rule out certain procedures, such as mere yes or no questions of roughly the form “Do you understand the relevant information?” on the grounds that they are simply far too simple to be used to indicate understanding of any degree. Some degree of discussion or questioning regarding the details of a decision will be required by the schematic solution, though it certainly seems plausible that different institutions will find themselves setting different standards, both for substantial understanding and for the detail of information the decider must express. The non-control condition, being slightly more public than the rest by its very nature, will have less interaction with the defense condition specified in the schematic solution. The schematic solution will presumably give a) a list of external conditions which the institution considers to be “unduly controlling” (and of course, different institutions will consider different forces to be undue influences), b) a list of internal conditions (e.g. mental illnesses) which would be considered unduly controlling, and c) a specified diagnosis procedure for how one would go about determining that a patient has an internal condition exerting undue control over them. By filling in all of the above blanks, an institution would have done a good deal towards ensuring that they were accounting for the concerns of Beauchamp and Childress’s definition of autonomous decision making.
WORKS CITED


New York, N.Y.: Oxford University Press.


