

ABSTRACT: 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.

### ***The Principle of Respect for Autonomy***

In *Principles of Biomedical Ethics*, Beauchamp and Childress propose a principle of “respect for autonomy” as a guiding principle of medical ethics. Focussing on autonomous decision making instead of autonomy as an overall property of agents, 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. Failing to meet any one of the conditions makes a decision non-autonomous. 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).

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 in terms of just an intended end because that is too vague. If a cancer patient, for example, 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 invasive surgery might play a part in achieving such an end, and by intending only “to get better”, the patient has not made sufficiently clear what they are after.

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, 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). Rational appeal is distinguished from appeals to emotion. 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). While vague, 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).

### ***Difficulties with the Principle of Respect for Autonomy***

In focusing on autonomous choice instead of the autonomy of agents, Beauchamp and Childress present 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"; Buss attributes such a view to Christman and Mele). Beauchamp and Childress are not focused on any feature of the agent's reasoning beyond the scope of a single decision. While their 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 a 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.

Beauchamp and Childress's criteria for autonomous decision making are plagued by a rule-following paradox, of the sort found in Kripke's version of Wittgenstein's private language argument. 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". The analogous problem for 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 the model 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? 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. As Wittgenstein states in §201 of the *Philosophical Investigations* “This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule.” (Wittgenstein §201) Taking the condition of understanding as an example, without clear public criteria for meeting the condition of understanding, the condition is 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<sup>1</sup> is created to meet the need of this public criteria must also be public in order to be sure that institutional policy reflects the spirit of the criteria for autonomous decision making.

A public definition of agent autonomy of the sort provided by Frankfurt, Christman, or Mele could serve as the criteria to determine if somebody had successfully met the conditions for autonomous decision. However, an advantage of Beauchamp and Childress’s account of autonomous decision making is that it is fairly open regarding: agent autonomy: conditions of intentionality, understanding, and non-control could be used in conjunction with numerous definitions of agent autonomy. The goal will be to find a way to incorporate public elements into Beauchamp and Childress’s criteria without incorporating a strict definition of agent autonomy. I propose that formal features of relational conceptions of autonomy provide a possible solution.

### ***Relational Autonomy***

---

<sup>1</sup> By “institutional policy” I mean the policies of any institution that would need to set standards for autonomous decision. Because Beauchamp and Childress’s work mostly concerns medical ethics, I have in mind medical institutions (hospitals, clinics, etc), though any institution that needs to judge whether a decision is autonomous may find itself in such a situation.

Relational theories of autonomy are those in which human relationships (which are necessarily public) are fundamental to the autonomy of an agent.<sup>2</sup> 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 to retain the openness of the principle of respect for autonomy.

A formal relational account of autonomy may be able to solve the problem of a lack of public criteria 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). I suggest 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. Her account of autonomy provides one answer to a notoriously tough case for many conceptions of autonomy: those of agents who have chosen life in a nominally community, e.g. the "happy slave." A substantive theory of relational autonomy (one that endorses a particular relationship or relationships as the source of autonomy) might declare that members of such a

---

<sup>2</sup> See the work of Marina Oshana, Jennifer Nedelsky, Catriona Mackenzie, and Natalie Stoljar for a variety of examples of and perspectives on relational autonomy. I cannot hope to touch on every, or even most, theories of relational autonomy here.

community lack autonomy (or 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 their choices, even if their choices are in line with values we might find “oppressive.” Westlund’s theory allows for the possibility that a member of a supposedly oppressive community could offer a suitably strong account of their choices, strong enough that an external critic could recognize the community member 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.<sup>3</sup>

### ***Combining Formal Relational Autonomy and the Principle of Respect for Autonomy***

What, then, is the result when one combines Beauchamp and Childress’s theory and Westlund’s theory? 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<sup>4</sup> 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 (but not sufficient) condition of autonomous decision making. If Westlund’s theory needs filling out, then Beauchamp and Childress’s conditions of autonomous decision making may fit the bill. Giving up on the three conditions as jointly sufficient is a small price to pay in order to salvage the three conditions.

---

<sup>3</sup> This looseness brings its own problems. There are numerous community values we generally find distasteful (e.g. extreme prejudice on the grounds of race, sex, or religion) and probably do not want to defend or use as the basis of a defense for various actions. For our purposes, though, that does not matter. We are trying to get an account of autonomous choice, not of the moral standing of any particular choice.

<sup>4</sup> For the sake of space, I will only discuss the conditions of understanding and intentionality.

In order to make use of Westlund's theory of relational autonomy, it needs some modification. Westlund problematically states her theory in terms of a disposition. If the theory is not somewhat modified, it will 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... 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)

Stated in terms of dispositions, Westlund's theory does nothing to resolve issues of privacy. The most straightforward solution is to find a way to state Westlund's theory without referring to private features of people, such as dispositions.

It is reasonable to recast Westlund's theory of autonomy in terms of *actually* defending the choices one makes. 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). Clearly, actually defending 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 not perfect. What counts as a "sufficient defense" is not entirely clear and may change depending on community values or institutional policy. One reasonable stipulation 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 and Beauchamp and Childress's positions.

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. A perfect understanding of the relevant information is unlikely to ever actually be achieved, but the standard of "substantial understanding" is reasonable. 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 and the possible payment plans for each treatment. These are the sorts of details one 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.

The defense condition cannot solve all of the issues surrounding the condition of understanding. While the defense condition might be used to set the 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 unlikely) that a patient could misconstrue relevant information but 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 could help indicate that a patient has a substantial

understanding, though it would not be foolproof. Nor would a suitable defense of a choice 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 would set a threshold for, and serve as an indicator of substantial understanding, assuming a decider *does* have a correct understanding, and that they have all of the relevant information.

The condition of intentionality could also be improved by incorporating the defense condition. The intentionality condition 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 by having a plan privately. This, however, leads to the problem discussed above: 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. Take the case of our hypothetical 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. In this case, with respect to ensuring that the patient’s decision was autonomous, there would be no difference between thinking they had a plan and actually having a plan. 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 the 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 and satisfied the condition of intentionality, 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 has a plan, as well as the values that underlie their 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 institutional policy. 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 has a plan. 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.

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. Imagine, then, that the institutions 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 it’s risks?”) or the mere signing of a form (“Sign here to indicate that you understand the procedure you are consenting to”). Such a practice does nothing to ensure that the condition of understanding is actually met. If we accept, as Kripkenstein 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. 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.

Kripke’s solution to the private language problem is 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)

Our 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 judges (e.g. physicians, nurses, hospital administrators) about whether a decider has met institutional criteria.

The defense condition places demands on the structure of institutional policy, but not the content. For example, 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. Of course, an institution could develop 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 make the understanding condition entirely trivial. In this case, the defense condition will be of no 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.

Beauchamp and Childress's definition of autonomous decision making incorporates three conditions: understanding, intentionality, and non-interference. Their definition of autonomous decision making is subject to a Kripkensteinian rule-following problem. One can escape from this by ensuring that the conditions of autonomous decision making are met publically. 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 publically. While the defense condition cannot solve all potential problems with Beauchamp and Childress's definition of autonomous decision making, I believe it provides a substantial improvement.

## Works Cited

- Beauchamp, Tom and Childress, James. *Principles of Biomedical Ethics, 7th ed.* Oxford University Press.
- Buss, Sarah, "Personal Autonomy", *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), URL <http://plato.stanford.edu/archives/win2014/entries/personal-autonomy/>.
- Christman, Jonathan. "Relational autonomy, Liberal Individualism, and the Social Constitution of Selves". *Philosophical Studies* 117 (1–2): 143–64.
- Kripke, Saul A. *Wittgenstein on Rules and Private Language: An Elementary Exposition.* Cambridge, Mass.: Harvard University Press, 1982.
- Westlund, Andrea. "Rethinking Relational Autonomy". *Hypatia*, 24: 26–49. 2009.