AUTONOMY AND INDOCTRINATION: WHY WE NEED AN EMOTIONAL CONDITION FOR AUTONOMOUS REASONING AND REFLECTIVE ENDORSEMENT*

By Mirja Pérez de Calleja

Abstract: *I argue that none of the main accounts of autonomy in the literature can explain the fact that people who undergo a certain subtle but powerful kind of indoctrination are not autonomous or self-governing in reflectively acquiring and endorsing the views, values, goals, and practical commitments that they are successfully indoctrinated to adopt. I suggest that, assuming there are historical conditions on autonomous reasoning and reflective endorsement, there is a condition that specifically concerns emotions: the person's emotional state and dispositions, and her web of emotional dependencies. I explain what we know so far about the kind of indoctrination on which I focus, and I motivate the claim that people who are successfully indoctrinated in this way are not self-governed in reflectively acquiring and endorsing even the first views and values that they adopt as a result of indoctrination. I argue that this heteronomy is not explained by any of the accounts that postulate historical conditions on autonomy: neither by classical accounts such as Rousseau's and Piaget's, nor by so-called historical accounts in the contemporary literature, nor by relational accounts. I argue that an accurate account of autonomy must include an emotional condition on autonomous reasoning and reflective endorsement that goes beyond the emotional conditions postulated or implied by historical accounts, and I offer a tentative sketch of this condition.*

KEY WORDS: doxastic autonomy, volitional autonomy, indoctrination, extremism

I. Indoctrination through the Internet

Before explaining the kind of indoctrination I am concerned with, let me start with some introductory remarks about autonomy and indoctrination. Autonomy is self-determination, self-governance, or self-rule. Heteronomy, in contrast, is being ruled by others or by something external to oneself. On the approach I will adopt, autonomy or self-rule consists in determining for oneself one's own views, values, deliberative strategies and policies, practical commitments, goals, and actions (insofar as this is possible), as opposed to having them set by others. On this approach, being an autonomous person, which involves choosing and acting autonomously on a regular basis, requires doxastic and volitional autonomy,

namely, autonomy in having the beliefs and pro-attitudes one has. Importantly, I assume that being doxastically and volitionally autonomous requires self-governance in the exercise of theoretical and practical reason by which one reflectively acquires, modifies, and rejects one's views, values, and so on.[1] (I do not assume that all autonomously held beliefs and pro-attitudes are reflectively acquired or reflectively endorsed, but only that an autonomous person has many reflectively acquired and endorsed beliefs and pro-attitudes.) My argument is aimed at theorists who also accept this, and hence postulate historical conditions on doxastic and volitional autonomy.

This essay does not address the difficult question of how children, who are initially not autonomous, become autonomous persons.[2] For the purposes of my argument, I assume that there is some way in which this happens. Thus, I leave aside the skeptical worry that there might be no principled difference between indoctrination and a good education. My goal is to convince theorists who accept a historical account of autonomy that we need to add emotional conditions on the reflective acquisition and endorsement of beliefs, pro-attitudes, policies, and commitments.

The Oxford English Dictionary defines indoctrination as "the process of teaching a person or group to accept a set of beliefs uncritically."[3] Indoctrination is a normative notion: it is a wrong thing to do and a bad thing to suffer, because it undermines autonomy. An extended approach in the philosophical literature is that a teaching technique undermines autonomy if it makes people accept, without proper critical scrutiny, doctrines that they should come to accept only through critical scrutiny.[4] On this general approach, critical scrutiny must meet formal rationality conditions (rules of deductive logic, probabilistic reasoning, and the like) if it is to ground doxastic and volitional autonomy, and it must be uninfluenced by deception, traumas, phobias, Pavlovian conditioning, and other kinds of manipulation. As I understand this condition, obtaining and keeping doxastic and volitional autonomy requires *self-governed* critical scrutiny in the reflective acquisition, endorsement, modification, and rejection of at least some key views, values, and so on.

There is no agreement on what specifically counts as indoctrination and what doesn't. For example, there is a long debate about whether religious

[1] This view is not uncommon. See, for instance, Michael Bratman, "Autonomy and Hierarchy," *Social Philosophy and Policy* 20, no. 2 (2003): 156–76, and Andrea C. Westlund, "Rethinking Relational Autonomy," *Hypatia* 24, no. 4 (2009): 26–49.

[2] Thanks to an anonymous referee for *Social Philosophy and Policy*, for an extensive and illuminating report that made me realize that I should leave this topic out of this essay.

[3] Available online at: https://en.oxforddictionaries.com/definition/indoctrination

[4] Note that, on this approach, not all doctrines are supposed to be accepted through critical examination. Learning mathematical and chemical doctrines from testimony, say just through repetition and without critical scrutiny, is not considered indoctrination, presumably because critical scrutiny should not be involved in learning these doctrines for the first time. Thanks to Stephen Kearns, Manolo Martínez, and Nick Wiltsher for pushing me to clarify this.

education is inherently indoctrinatory. Unlike behavioral conditioning, which *bypasses* the person's use of reason, indoctrination involves an unavoidable amount of reasoning on the part of the person who is being manipulated. Indoctrination instills beliefs about how things are or should be, and also purported justifications for these beliefs, which the person must reason through to internalize. The problem is that this reasoning is not self-governed (whatever exactly this amounts to), but rather guided through narrow paths by the indoctrinators.

This essay focuses on the kind of indoctrination that ISIS uses to recruit teenagers and youngsters in Western countries. It is not an easy task to gather empirical data about how terrorist groups use indoctrination. But researchers agree that there is no profile shared by most people who get indoctrinated into violent extremism, apart from a young age and male gender. (In the particular case of ISIS, only a young age.) There is diversity in recruits' social class, ethnicity, family's immigration history, education, religion, personality profile, and psychopathological disorders. Most recruits are mentally healthy from a clinical perspective. It is not the case that most suicide bombers antecedently had suicidal tendencies. And social networks and environments do seem to be crucial in attracting youngsters to terrorist movements.[5]

The kind of indoctrination that ISIS uses to convert Westerners into jihadism happens exclusively through the Internet (with no physical contact between the people involved), but it is not limited to deception. It does not work on everyone, but it has powerful effects on some people. We don't know whether any criteria (apart from age) are used to target the youngsters, nor whether there are psychological or other preconditions that explain why this indoctrination works on certain people but not on others.

These are, in rough terms, the steps that this indoctrination takes. First, an ISIS member who has been trained to radicalize and recruit Westerners contacts the person through social media and starts casual conversations, pretending to be someone with similar interests or experiences. At first, these conversations are not political or religious at all. Once an affinity is established, the conversations move to private messaging, and become more and more centered on Islam (or rather, ISIS's version of it) and jihadism. The conversations are very frequent, and soon they become almost continuous. The targets are pointed to websites that explain and justify jihadism, and to macabre and visually impacting videos that supposedly illustrate some of the lies told, and exhort people to, for instance, move to Syria to

---

[5] Alex Schmid provides the most thorough review I know of the empirical literature on this subject. See Alex P. Schmid, "Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review," *ICCT – The Hague Research Papers* (The Hague: International Centre for Counter-Terrorism, 2013): 26. See also Dounia Bouzar, "A Novel Motivation-based Conceptual Framework for Disengagement and De-radicalization Programs," *Sociology and Anthropology* 5, no. 8 (2017): 600–602.

fight for justice or to save babies from dying at the hands of evil powers. Once targets show sufficient conviction, they enter online "support groups" formed of converts who message each other constantly, first with reassuring messages about their cause and expressions of companionship and support, and eventually with collective plans to join ISIS militant groups in Syria or Iraq.

I will focus on Dounia Bouzar's work with youngsters that were radicalized by ISIS, because she had access to more and better-quality testimony of radicalized and ex-radicalized people than any other researcher that I know. Thanks to the French government's support, Bouzar held therapy sessions with over a thousand radicalized and ex-radicalized youngsters, many of whom were in preventive house arrest at the start of therapy. The sessions were held for long periods of time, and with the regular participation of family members and of "repented" or reformed jihadists. I take this to be a clear improvement on sporadic interviews carried out in prisons, so far the most common way for researchers to gather data about radicalized people's perspective on their own radicalization.

Seeing clear patterns in the self-reports and Internet activity records of these youngsters, Bouzar identifies seven "primary motives" or "myths"[6] leading a progressive turn to extremism. All of them fall under one of two types: "the search for an ideal, be it an ideal self, world, partner, or community; [. . .] or the escape from the real world towards 'another place' that is supposed to be better."[7] Bouzar and her team found ample evidence that indoctrinators adapt their conversations, pictures, and videos to the person's particular motives to be potentially interested in Islam and jihadism. Indoctrinators also take into account the function that the person would have in ISIS: the messages that are useful to move someone to be an ISIS bride are significantly different from the ones used to recruit an ISIS fighter. As I mentioned, this tailoring seems to be made roughly according to seven general motives, or "myths" as Bouzar calls them, to which she gives rather self-explanatory names: "the quest for a better world";[8] "the Mother Theresa myth";[9] "the savior myth";[10] "the myth of the sleeping beauty";[11] "the Lancelot myth";[12] "the myth of Zeus";[13] and "the myth of the fortress."[14]

Conspiracy theories were used in all of the cases that Bouzar studied, and she compellingly speculates that jihadists specialized in online

---

[6] Bouzar, "A Novel Motivation-based Conceptual Framework for Disengagement and De-radicalization Programs," 606.
[7] Ibid., 606.
[8] Ibid., 606–7.
[9] Ibid., 607.
[10] Ibid., 607–8.
[11] Ibid., 608.
[12] Ibid., 608–9.
[13] Ibid., 609.
[14] Ibid., 609.

indoctrination use conspiracy theories to create mistrust for authorities, family, and friends, in their effort to destabilize the person emotionally.[15] A growing mistrust of friends and family is accompanied by a growing emotional distance from them, and by a growing emotional dependence on indoctrinators and other indoctrinated jihadists. In sum, daily anxiety, stress, fear, mistrust, and suspicion are induced in order to make the youngsters seek safety in the indoctrinators.[16] In particular, according to Bouzar, the fear of being persecuted is fundamental in this process.[17]

This emotional manipulation is far more extreme than the emotional manipulation that, arguably, any moral education must use in order to be effective. I am ready to grant, in particular, that an effective moral education occasionally needs to induce negative emotions such as fear and guilt. But a good moral education does not use extreme and constant anxiety, fear, guilt, paranoia, and so on.[18]

Relatedly, these emotional influences on thought are much stronger and more constraining than the unendorsed emotional motives that often lead people to deceive and blind themselves in theoretical and practical reasoning, such as love, jealousy, pride, envy, and so on. Even if unconscious emotional constraints that the person would not endorse as reasons to think or do things do not undermine autonomy of thought, the emotional constraints orchestrated by ISIS indoctrinators may well undermine self-governed reasoning.

Consider one of the testimonies discussed by Bouzar, which is very representative of the testimonies that she and her team studied. When she was 16, Norah, born and raised in France and with no extremist friends or relatives, attempted to go to Syria twice to join ISIS. Years afterward, during her therapy, she described in this way the constant fear, paranoia and anguish that she felt as she was being indoctrinated to travel to Syria to be an ISIS bride:

> It was obvious: our enemies were everywhere. Since we knew the truth, we were constantly under surveillance. And the more we felt under surveillance, the stronger became our conviction of knowing the truth. I saw this as belonging to an authentic group, in which we were the most "awakened" ones. We were being tracked because they wanted us asleep and indoctrinated . . . I was afraid that people were trying to lull me to sleep and to hurt me. I had to stay awake, no matter what.[19]

It is remarkable that Norah's explicit fear of being indoctrinated helped indoctrinators do their job. Norah had an explicit commitment to knowing

---

[15] Ibid., 603–4.
[16] Ibid., 603–5.
[17] Ibid., 604.
[18] Thanks to Stephen Kearns and Michael McKenna for pushing me to clarify this point.
[19] Cited by Bouzar, "A Novel Motivation-based Conceptual Framework for Disengagement and De-radicalization Programs," 604.

the truths that she thought evil powers want to keep hidden from people. She was instilled with a sense of duty to avoid the indoctrination to which everyone else falls prey: she learned that she was chosen by God to see what others don't want to see. She was convinced that she was one of the few who can save the sleeping masses from the damnation they would otherwise be doomed to. At the early stages of her conversion, she had rather intricate epistemological discussions about her conspiracy theories with her school friends. All of these things are very common among the cases that Bouzar and her team studied.

One of Bouzar's conclusions is that an effective indoctrination requires that an emotional relation be established between indoctrinator and indoctrinated person, and similar conclusions have been drawn by empirical studies of indoctrinated terrorists that were made in other countries with different methods and different sources of testimony.[20] Bouzar[21] suggests that what explains how these formerly normal youngsters converted to jihadism through online conversations and exposure to propaganda is in part that they were looking for bonding or for identification with a network of people, and in part that they were progressively more and more convinced and concerned by the ideology of ISIS.

Thus, according to Bouzar (and, in earlier work, Bouzar and Martin[22]), the emotional aspect of indoctrination is essential. In particular, a need for bonding, intimacy, compassion, or closeness seem to be key in most cases.[23] And just as the emotional influence of indoctrinators is crucial for effective indoctrination, in order to reverse the effects of indoctrination it is crucial to approach the problem from an emotional perspective. Indeed, an emotional approach is the first and most important step of the therapy advocated by Bouzar and Martin. Their job, as they see it, is to help victims *de-radicalize* themselves, mostly through psychoanalysis, and in particular by making victims confront the particular hardships that led them to embrace jihadism as their new ideology and way of life.[24] Their therapy addresses the person's emotional dependence on the indoctrinators that she bonded with, and it works mainly by helping the person deconstruct and criticize the paranoia and extremist worldview that she acquired, and by offering alternative worldviews that do not provoke constant anxiety and fear and a sense of lack of safety. On this model of a *de-radicalization* therapy, treating the emotions caused

[20] See Schmid, "Radicalisation, De-Radicalisation, Counter-Radicalisation."

[21] Bouzar, "A Novel Motivation-based Conceptual Framework for Disengagement and De-radicalization Programs," 602.

[22] Dounia Bouzar and Marie Martin, "Méthode Expérimental de Déradicalisation: Quelles Strategies Emotionelles et Cognitives?" *Pouvoirs* 158 (2016): 83–96.

[23] Bouzar, "A Novel Motivation-based Conceptual Framework for Disengagement and De-radicalization Programs," 602–3.

[24] Ibid.

by the ideology is inseparable in practice from intellectually deconstructing and criticizing the ideology.[25]

In the early stages of their conversion, self-recruited jihadists like Norah meet all formal rationality requirements, exhaustively gather information, critically consider alternative explanations of phenomena, daringly challenge social expectations, abandon views and values that they used to take for granted and identify with, and adopt views and values considered abhorrent by virtually everyone in their community (including, formerly, themselves). In sum, they reflectively endorse their new views, values, and practical commitments through eager, truth-driven scrutiny. Indeed, they are explicitly asked to think by themselves and challenge common prejudices and illusions, and this is precisely what they think they are doing as they enter deeper and deeper into the worldview and way of life that ISIS wants for them.[26] To clarify, I do not think that the methods used by ISIS undermine autonomy because of the recruiters' intention to instrumentally manipulate their targets, nor because of their readiness to deceive whenever this suits their purposes. In my view, even if these conversion processes had been caused by a sophisticated computer program whose operation was accidentally triggered, and all the conversations involved had been held with advanced AI software, the affected subjects would still have had their autonomy undermined, independently of whether that should count as indoctrination.

These subjects' reasoning and reflective endorsements, in the early stages of conversion, seem to satisfy all formal rationality requirements (such as rules of deductive logic and probabilistic and instrumental reasoning), and meet all of the historical conditions on doxastic and volitional autonomy that have been identified in the literature. However, I submit that, as a consequence of the powerful emotional manipulation to which they are subjected, their theoretical and practical reasoning about questions related to jihadism plausibly stops being self-governed early on in their conversion process, and consequently their reflective acquisitions and endorsements of new views and values are heteronomous. I suggest that these subjects' reasoning stops being self-governed as their constant state of alert, fear,

---

[25] See also Bouzar and Martin, "Methode de Deradicalisation: Quelles Stragegies Emotionelles et Cognitives?"

[26] Despite all this reasoning on the part of converts, it seems natural to refer to the techniques used by ISIS as indoctrination, and they are indeed called "indoctrination" in studies by governments, armies, and counterterrorism centers around the world. Of course, this doesn't settle the question of whether or not these techniques count as indoctrination in some specific philosophical sense that might be defined, or the question of whether or not these techniques, when successful, undermine autonomy. While I do claim that these techniques undermine autonomy, I do not claim that the fact that these youngsters are not autonomous in reflectively endorsing the relevant views, values, and commitments exempts them from blame for having these and being guided to action by them. I set aside in this essay the delicate question of their moral responsibility. This is why my discussion of the examples leaves aside several aspects of the agents' histories and present conditions that may be relevant to determine their moral responsibility.

paranoia, guilt, and anguish starts blocking salient paths of conscious reasoning that might end up challenging the jihadist ideology. One might say that they are *incapacitated*, or perhaps psychologically unable, to reason in certain obvious alternative ways in which they *should* be able to reason, in the sense that they not only have the conceptual resources and emotional nuances necessary to reason in these ways, but also motives and internal and external reasons. But it becomes increasingly unbearable for the subjects to challenge the jihadist ideology, because they become increasingly invested in urgently repairing terrible evils by being good jihadists.

This kind of *incapacitation*, if we may call it that, seems to be the kind of thing that can obstruct self-government in theoretical and practical reasoning. Much like people with severe unconscious traumas, it is very plausible that abstaining from reasoning in certain ways becomes compulsive for these people. Thus, the ways in which they actually reason, taken in isolation, do not fail to meet any rationality conditions for autonomy; but the fact that they never reason in other salient ways betrays the fact that their reasoning is not self-governed after all.

Consider the emotional state they are in, and the set of emotional dispositions and web of emotional dependencies that they acquire early on in the indoctrination process. They are terribly anguished by the thought that they might really be in a unique position to redeem others' evil through a jihadist way of life (that they were chosen by God to save their own family from eternal damnation, to save innocent victims of war, and so on). This plausibly brings to them something that can be aptly described as an obsessive compulsion to fulfill their moral duty and be "good jihadists." They increasingly feel they are potentially under surveillance, and in danger of being discovered and imprisoned by the authorities. This probably increases their subjective certainty of the truth and value of jihadism. Moreover, they are in constant communication with jihadists to whom they feel they must justify their views, values, and actions.

Once these processes start, it is only to be expected that they escalate and end up in dogmatism. But, in the early stages, subjects seem to meet all of the conditions for autonomous thought and reflective endorsement that different accounts of autonomy identify. I suggest that, if unconscious phobias and traumas undermine self-government of thought, so does this kind of indoctrination, from its early stages on. Thus, I suggest that there are emotional conditions on self-government of theoretical and practical reasoning that have not been identified in the literature, which these indoctrinated subjects fail to meet.[27]

---

[27] I do not believe, though, that the ideal of autonomy includes an emotional independence from others, nor an emotional detachment from the doctrines one believes and stands by.

II. Classical Accounts of Autonomy

Classical accounts of autonomy such as Rousseau's[28] and Piaget's[29] do justice to the fact that, even if one willingly governs oneself by a certain set of views, goals, and principles, one is not really the author and master of one's life if these have been instilled in one, as opposed to one's having reflectively acquired and endorsed them through honest, truth-driven critical scrutiny. The Rousseau-Piaget view of education emphasizes the value of fostering the child's independent thinking about all matters, as opposed to her mindless swallowing of the socially accepted views and ways of life. In particular, doxastic autonomy (that is, autonomy in believing what one believes) requires the person's acquisition and maintenance of beliefs to depend on the conclusions she reaches in conscious critical scrutiny.

It seems clear that youngsters who get indoctrinated in the way described above do meet classical conditions for doxastic autonomy (charitably interpreted so that they are not impossible to meet) in the early stages of their conversion process, when they haven't yet become dogmatic about their new worldview. However, their emotional conditions in these early stages of the process seem to undermine their self-government in reasoning about questions related to jihadism, and in reflectively acquiring and endorsing new views and values related to these questions.

Someone might object that these subjects do not meet classical cognitive independence conditions on self-government, because their thought, while perhaps independent from mainstream influence, is not independent from strong influence.[30] I agree that this is plausibly what is going on at the last stages of conversion and afterward. But it does not seem psychologically plausible to say that a person can passively, without honest critical scrutiny and just *driven* by the influence of others, abandon central views and values that one has always taken for granted, and adopt radically different ones. Moreover, in light of what we know about the kind of indoctrination discussed, it does not seem accurate to describe the process as one in which the subject adopts whatever views and values certain people manifest, just because they are the ones manifested by these people. It is more plausible that subjects come to invest certain sources with authority as a consequence of their gradual acceptance of the relevant doctrines, rather than as a cause of this acceptance. What happens in the early stages of the indoctrination process cannot be explained by invoking a reason-overriding influence of others.

---

[28] See Jean-Jacques Rousseau, *Emile, or Education* (London and Toronto: J. M. Dent and Sons, 1921), originally published as *Émile, ou de l'Éducation* (Paris: Jean Néaulme, 1762).

[29] See Jean Piaget, *Science of Education and the Psychology of the Child* (New York: Penguin Books, 1971), originally published in 1969.

[30] Thanks to Stephen Kearns for raising this objection.

Classical accounts of autonomy understand self-governance as something that is perfectly achieved only by a cognitively and motivationally independent person, namely one whose views, desires, values, goals, principles, and so on do not depend on other people's, and are instead fully determined by the person's own experiences and critical scrutiny. Contemporary accounts of autonomy do not make this implausible claim.

III. Contemporary *Historical* Accounts of Autonomy

Unlike classical accounts of autonomy, contemporary accounts acknowledge, explicitly or implicitly, that people are unavoidably shaped by their formative education and upbringing, and that nobody is cognitively and motivationally independent. So-called *historical accounts* of autonomy[31] claim that an agent acts freely (and not just intentionally) only if the psychological constitution that motivates her choices and actions has a certain kind of history.[32] In particular, even if one endorses one's goals, practical commitments, and choices, one is not autonomous if this endorsement was conditioned by one's caretakers and educators when one was an infant, or was brainwashed into one by powerful neuroscientists during one's sleep, or the like. One becomes an autonomous person only if one acquires and maintains one's views, values, desires, goals, and so on as a result of the right kinds of processes, and this involves most notably an absence of effective manipulation.[33]

---

[31] Such as John M. Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge, UK: Cambridge University Press, 1998), Ishtiyaque Haji, *Moral Appraisability: Puzzles, Proposals, and Perplexities* (New York and Oxford: Oxford University Press, 1998), Ishtiyaque Haji and Stefaan E. Cuypers, *Moral Responsibility, Authenticity, and Education* (Routledge International Studies in the Philosophy of Education, Taylor and Francis, Kindle Edition, 2008), Robert Kane, *The Significance of Free Will* (New York and Oxford: Oxford University Press, 1996), and Alfred Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York and Oxford: Oxford University Press, 1995).

[32] As I mentioned, I am leaving aside so-called *nonhistorical* accounts (such as Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy* 68, no. 1 [1971]: 5–20, and Gary Watson, "Free Agency," *Journal of Philosophy* 72 [1975]: 205–20), which maintain that the formative history of a person's psychological make-up is irrelevant to her present capacity for autonomous thought and action. Nonhistorical accounts imply that a person can choose and act autonomously driven by motives that were instilled in her through indoctrination, brainwashing, or any other kind of manipulation. Thus, these accounts do not rule out the indoctrination that concerns me, but they welcome the result that its victims are autonomous. As I explained, my argument is aimed at theorists who assume, like I do, that there are historical conditions on autonomy. Interestingly, though, the conditions of nonhistorical accounts do not rule out either the intuitively problematic way in which, as a result of the kind of indoctrination I am concerned with, emotional needs plausibly come to limit the possible results of particular episodes of reasoning, self-understanding, and practical deliberation, including the endorsement and rejection of desires as desires on which it would be good to act. And this might be considered a problem by nonhistorical theorists.

[33] In addition, many incompatibilists (i.e., philosophers who claim that free will and physical determinism are incompatible) have argued that, for somebody to be morally responsible for having and acting on their beliefs, desires, goals, values, and so on, critical reflection must have *made a difference* to their present attitudes and tendencies, on an indeterministic reading of this expression. The question whether these incompatibilists are right is beyond the scope of this essay.

The kinds of manipulation that these accounts rule out are unlike ordinary forms of indoctrination in that they completely bypass the victim's critical thinking. These forms of manipulation include *implanting* values and goals by manipulating a person's brain during sleep,[34] synchronically controlling a subject's mental and actional processes through electrical stimulation of the brain,[35] *beating into* a child the disposition to act morally,[36] instilling such habits and dispositions via *shock therapy*,[37] employing the kind of character-determining clandestine conditioning featured in B. F. Skinner's novel *Walden Two*,[38] and subjecting kidnapped children to Pavlovian conditioning that sometimes succeeds at instilling the deliberative habit of prioritizing the leader's interests.[39] These rather science-fictional ways to manipulate a person's acquisition of views, values, goals, deliberative strategies, and so on are not relevant to the indoctrination problem at hand, because they completely bypass victims' critical thinking, ensuring that they fail to meet the rational conditions on autonomy that the indoctrinated people who concern me do seem to meet.

On the other hand, while most historical accounts incorporate what one might call a *normality condition* on the emotional make-up of autonomous agents (roughly, an absence of pathologies such as psychopathy, autism, and the like), such a condition does not rule out the emotional states and dependencies of people who are effectively subjected to the kind of indoctrination that this essay discusses, at least during the early stages of the process. If this indoctrination undermines doxastic and volitional autonomy in the early stages of conversion, then contemporary historical accounts of autonomy fail to identify sufficient conditions on the autonomous reflective acquisition and endorsement of beliefs, pro-attitudes, and so on.

---

[34] See Mele, *Autonomous Agents: From Self-Control to Autonomy*, Haji, *Moral Appraisability: Puzzles, Proposals, and Perplexities*, and Haji and Cuypers, *Moral Responsibility, Authenticity, and Education*.

[35] See Mele, *Autonomous Agents*, and Haji, *Moral Appraisability*.

[36] See Haji, *Moral Appraisability*.

[37] See ibid.

[38] See Kane, *The Significance of Free Will*, and Haji and Cuypers, *Moral Responsibility, Authenticity, and Education*.

[39] See Haji, *Moral Appraisability*. As it turns out, Haji (120–21) argues that this kind of conditioning does not undermine autonomy, as long as it "leaves untouched [a successfully conditioned victim's] capacities to become aware of his new deliberative habit, to reflect critically on it, and to counter its influence" (120). I do not share the intuitions about thought experiments on which Haji relies, and consequently I am not convinced by his argument. But in any case, as Haji points out, the techniques featured in this thought experiment "bypass [victims'] capacities for cognitive control over their own mental lives" (120). Successfully manipulated children acquire the target deliberative habits through conditioning, and not by assessing and endorsing the priorities they acquire.

IV. Relational Accounts of Autonomy

So-called *relational accounts of autonomy* emphasize that being an autonomous person does not mean determining one's way of life independently from other people, since any person is formed and lives within a specific social structure and web of social relations. In particular, any person's identity is partially shaped by social determinants such as race, ethnicity, class, and gender.[40] The social conditions that different relational accounts require for autonomy are diverse. But a unifying worry is that oppressive social conditions may undermine the autonomy of people who seem, at first sight, to determine their own way of life.

The historical accounts discussed in the previous section are far from claiming that autonomy requires cognitive and emotional independence from others, but they do not address the social dimension of autonomy, nor, in particular, the question of what social dependencies and relations might undermine autonomy. Hence, they are not relational accounts in the sense just defined,[41] though they are not in tension with incorporating social conditions on autonomy. Despite their explicit consideration of social relations and dependencies, nor do relational accounts rule out the kind of indoctrination this essay discusses. In particular, these youngsters' conscious exercise of critical reflection and desire endorsement, and their resulting acceptance of the views, values, and commitments which I claim they are not autonomous in having, are not characteristically constrained by oppressive ideologies and roles that have been internalized since childhood and are not directly consciously accessible.

Let me briefly consider Andrea Westlund's relational account, which I initially took to be able to explain why the indoctrinated youngsters I am concerned with are not autonomous in reflectively acquiring and endorsing their jihadist views, values, and commitments. For Westlund, autonomous choice and action requires self-governance of practical reasoning,[42] which in turn requires the disposition to *hold oneself answerable* to external critical perspectives for one's action-guiding commitments, such that one is disposed to provide one's justifications for endorsing certain desires and values and for treating them as justifying reasons for

---

[40] For an influential definition of relational accounts along these lines, see Catriona MacKenzie and Natalie Stoljar, "Introduction: Autonomy Refigured," in Catriona MacKenzie and Natalie Stoljar, eds., *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self* (New York: Oxford University Press, 2000), 4.

[41] Haji and Cuypers, in *Moral Responsibility, Authenticity, and Education,* say that they defend a "relational view of authenticity" (position 162), where "authenticity" means roughly what I mean by "autonomy" (see chap. 5). However, by "relational" they do not mean what I do, but rather, as they put it, "forward-looking" (position 688). They write: " . . . we defend a relational view of authenticity according to which motivational (and other) springs of action are authentic or inauthentic only relative to whether later behavior that issues from these springs is behavior for which its agent is responsible" (positions 162–63).

[42] Andrea C. Westlund, "Rethinking Relational Autonomy," *Hypatia* 24, no. 4 (2009): 31.

action, in the face of legitimate critical challenges from real or imagined people with a different point of view.[43] Lacking this disposition renders one's practical reasoning disconnected from and insensitive to external justificatory pressures.[44] And, Westlund writes, "[b]eing impervious to critical challenge in this way is an excellent candidate for what it is to be gripped by an action-guiding commitment or bit of practical reasoning as opposed to governing it, which is precisely the distinction of which we need our account of autonomy to make sense."[45]

Westlund discusses how this applies to what she calls "deeply deferential agents," namely "those who endorse their deference but have no basis for doing so that is not itself deferential."[46] With these persons, it doesn't seem possible to genuinely engage in critical dialogue about the action-guiding commitments and self-governing policies[47] they endorse, since they merely keep pointing to certain others' justifications for these.[48] Westlund suggests that deeply deferential agents (and some depressed agents) are plausibly *in the grip of* the concerns that motivate their known unchallenged reasoning-governing policies—in the autonomy-undermining way in which one can be *in the grip of* reflectively endorsed desires as opposed to governing one's endorsement of them[49]—in such a way that these policies are not *their own* in the sense required for autonomy.[50]

Does this explain the heteronomy of the youngsters I am concerned with at the early stages of their indoctrination process? Are they deeply deferential toward their indoctrinators? I do not think so. The problem is that these subjects do hold themselves answerable to critical interlocutors for their views and commitments, and they do think (and often insist) that they are "answering for themselves"[51] when they respond to critical challenges. They certainly do not *show* deference to certain others when they respond to challenges for justification. Neither their upbringing nor their social conditions cornered them into accepting their new views and commitments, and I cannot see how they could have passively drifted into deferring to certain others in justifying their views, values, and commitments. Their conversion involved accepting authority sources that were not imposed on them, and rejecting sources whose authority was taken for granted in their communities.

---

[43] Ibid., 33–40.

[44] Ibid., 34.

[45] Ibid., 34.

[46] Ibid., 32.

[47] Westlund uses *self-governing policies* in Bratman's sense, namely personal policies concerning what weight to give to certain considerations in practical reasoning about one's conduct. (See Michael Bratman, "Autonomy and Hierarchy," *Social Philosophy and Policy* 20, no. 2 [2003]: 156–76.)

[48] Westlund, "Rethinking Relational Autonomy," 33.

[49] Ibid., 31–33.

[50] Ibid., 32–33.

[51] Ibid., 34.

Thus, these subjects do seem to meet, at the early stages of their conversion process, Westlund's condition of answerability to critical challenge, as she formulates it. For instance, she writes: "One who is disposed to hold herself answerable to others treats her commitments as something for which she herself is interpersonally accountable: they are neither simply brute facts about her, nor, ultimately, assignable to anyone else. She purports to speak on her own behalf, or to ''represent'' herself in interpersonal dialogue."[52] This seems to apply squarely to the indoctrinated youngsters I am concerned with, given the information we have about them. The same is true of this statement: "It is precisely insofar as one is responsive to perspectives that are *not one's own* that one demonstrates that one is not simply in the grip of one's own commitments, but responsive to normative pressures to which those commitments are subject."[53] Given that virtually all of the youngsters that ISIS indoctrinates through the Internet come from backgrounds where jihadism is seen as abhorrent, in almost all cases of effective indoctrination, responding to the indoctrinators' critical challenges requires precisely responding to perspectives that are very alien to one's own. Lastly, consider this claim: "What marks an agent out as self-answerable is how receptive she is to the critical perspectives of others. The autonomous agent experiences those challenges . . . as having normative standing in her deliberations, and reacts as though she owes a response."[54] In the early stages of the process, subjects are deeply ambivalent about what side they are on, and they take justificatory challenges from both jihadist and anti-jihadist perspectives as having normative standing.

Once they have come to fully embrace their new ideology and lifestyle, these subjects do seem to lose responsiveness to critical challenge. The more convinced they are of their new worldview, the more fervently they defend it and define themselves by it, and the less receptive they are to critical challenges. They come to believe that some people are so lost and scared that they would never understand the truth, so that critical dialogue is a waste of time. But these things characterize fanatics in general, and they do not help identify what undermines self-government in the distinctive ways in which these subjects reflectively acquire and endorse the first views and values they adopt as a result of indoctrination. In the early stages of the indoctrination process, these youngsters are able and willing to engage in sustained debates with people with different perspectives who challenge them to justify their views.

At this point, the reader might worry that what really drives my intuition that the youngsters this essay discusses do not autonomously acquire their jihadist views and values is simply that I consider these to be abhorrent.

[52] Ibid., 35.
[53] Ibid., 36, emphasis in the original.
[54] Ibid., 37.

But I do believe that a person can autonomously come to endorse abhorrent views, values, and practical commitments.[55] Let me elaborate on this briefly.

Accounts of autonomy can be either *procedural* or *substantive*.[56] *Procedural* accounts impose merely formal conditions on autonomous agents' psychological make-up: capacities to reason well, to revise beliefs and goals in light of reasoning, and to reflectively endorse action-guiding desires, values, principles, and commitments. Procedural accounts do not impose limits on the contents of the beliefs, values, commitments, and so on, that a person can autonomously have, reason with, or act on. For instance, an addict who wholeheartedly endorses her desire for a drug as the desire that moves her to action may autonomously choose to use the drug even when her desire is irresistible.[57] And a woman who willingly lives subordinated to the men in her community may autonomously do so, if she is disposed to respond to legitimate external critical challenges by offering the reasons why she endorses her action-guiding commitments —as opposed to being disposed to refuse to engage in dialogue with real or imagined interlocutors with a different point of view, or disposed to just defer to certain others when asked to justify her commitments.[58]

In contrast, *substantive* accounts, besides imposing formal conditions on autonomous agents' psychology, also impose constraints on the contents of the views, preferences, goals, and so on, that can in principle be autonomously adopted and maintained. Some substantive accounts are *weakly substantive*: they impose only minimal constraints on the contents of autonomously held cognitive and conative attitudes. *Minimal* constraints in this sense may amount to an absence of severe mental illness and large-scale manipulation and delusion,[59] or to this together with an explicit or implicit acceptance of one's own worth as a critical thinker.[60] A *strongly substantive* account of autonomy goes further than this, and claims that there are substantial normative constraints on what an agent can autonomously take to be true, desirable, valuable, and worth-pursuing.

Someone might think that, even if procedural and weakly substantive accounts of autonomy cannot account for what goes wrong in the cases that worry me in this essay, perhaps a strongly substantive account would be able to do this. My problem with this solution is that I take strongly

---

[55] Thanks to Michael McKenna for pushing me to clarify this.

[56] I thank Randolph Clarke, Stephen Kearns, and Michael McKenna for criticisms on a previous draft that motivated including this distinction and the paragraphs that follow.

[57] See Harry Frankfurt, *Necessity, Volition, and Love* (Cambridge, UK: Cambridge University Press, 1999) for Frankfurt's last version of this account.

[58] See Westlund, "Rethinking Relational Autonomy," 26–49.

[59] See, for instance: Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*; Haji, *Moral Appraisability*; and Mele, *Autonomous Agents*.

[60] See, for instance, Paul Benson, "Taking Ownership: Authority and Choice in Autonomous Agency," in J. Christman and J. Anderson, eds., *Autonomy and the Challenges to Liberalism: New Essays* (Cambridge, UK: Cambridge University Press, 2005), 101–126, and Carolyn McLeod, *Self-Trust and Reproductive Autonomy*, (Cambridge, MA: MIT Press, 2002).

substantive accounts to be very implausible. First, someone may acquire bad values and commitments through normal processes, and we should count such a person as autonomous (assuming mental sanity, suitable critical scrutiny, and so on), if we want to keep the distinction between autonomy and moral goodness. I cannot argue here for the need of this distinction. But it I take it that the burden is on the strongly substantive theorist to argue that a person cannot become and be evil autonomously. Secondly, the kind of indoctrination I discussed (and other kinds too) could in principle be used to instill true views and good values. We can imagine an extreme scenario in which using this kind of indoctrination is justified because it is the only reliable way to get some people to adopt a certain way of life within a given period of time, and because if this doesn't happen these people will suffer terrible consequences that are far worse than a lack of autonomy and the distress involved in being indoctrinated in this way. We can also imagine, if necessary, that this lack of autonomy would only be temporary. We can make the scenario such that, after a period of time, people's lifestyle will no longer have potentially terrible consequences, and when that time comes the indoctrinators will reveal their methods to the subjects, who will eventually succeed at reversing the indoctrination, autonomously maintaining or modifying their views, values, and way of life.

Thus, even though I do assume that autonomy is valuable, and even if this view implies that there are weakly substantive constraints on the ways in which one can autonomously live, I am claiming that the lack of autonomy of the indoctrinated youngsters I am concerned with has a procedural or formal ground. In particular, what undermines their autonomy is not that they adopt false views and bad values, but that indoctrination impacts them in ways that impede their ability to critically scrutinize their views and values in a way that is suitable to ground autonomy of thought.

V. Motivating the Need for New Emotional Conditions on Autonomy-Preserving Reflective Endorsement

Given that the conversions I am concerned with meet all of the conditions imposed by so many accounts of autonomy in the literature, might not these conversions be self-governed after all? Perhaps we (or rather, I) should just give up the intuition that they are not self-governed. After all, these conversion processes, at least in their early stages, do not seem to involve insufficient reasoning, nor formally negligent reasoning, nor merely deferring reasoning. They involve both intellectual humility and intellectual courage, and a commitment to finding the truth sufficiently strong to make these converted youth ready to become monsters in the eyes of the community. However, I still do not think that these conversions are self-governed.

First, even though, admittedly, my claim that these youngsters are indoctrinated and heteronomous rests partly on my own intuitions, this impression is shared by experienced therapists, caring and knowledgeable family members and friends, and the very people who succeed in getting out of the grip of the extremist views and values to which they once converted. Their intuitions are formed from different perspectives, and after for years dealing daily with the processes of conversion. They all agree that the converted self was unrecognizable as the person before the conversion, and had lost normal psychological continuity with previous selves. The experience of the de-radicalized, as they describe it in therapy, is not just an experience of reflectively rejecting abhorrent views and values that one once reflectively and wholeheartedly endorsed, but rather an experience of struggling to keep oneself detached from an asphyxiating set of theoretical and practical commitments in which one used to feel unavoidably trapped. While these considerations are far from decisive, I believe that they carry some weight.

Second, one thing that I think should make us suspect that the reflective acquisitions and endorsements I am discussing are not autonomy-preserving is that the subjects come to adopt what were for them morally abhorrent commitments and goals, relatively fast, and without any revelatory first-personal experiences prompting them to re-evaluate their whole worldview and value system. It's not like they discover one day that there is a wall instead of the sky in the horizon (like in the movie *The Truman Show*), or that they undergo some transformative experience[61] that gives them new access to certain facts. Far from smoothly adopting a jihadist worldview and set of commitments through consistent critical scrutiny, they must first be eased into accepting views in the ballpark (a set of doctrines falsely identified as Islamic by the indoctrinators, conspiracy theories, and lies about current states of affairs in distant countries), and into specific rules of conduct that gradually depart from their old ones.

Finally, there are important analogies between the emotional aspects of the indoctrination I am concerned with and the emotional states associated with severe traumas—emotional states which are rather uncontroversially thought to undermine self-government in theoretical and practical reasoning. Arguably, severe traumas (and perhaps also some phobias and addictions) show that self-government of thought is not only a matter of how one actually reasons and reflectively endorses beliefs, pro-attitudes, and practical commitments, but also a matter of what other paths of reasoning one was psychologically able to go through instead, and what other reflective endorsements one might have made. As I explained in Section I, severe unconscious traumas can impose unendorsed emotional constraints on thought, by blocking certain salient reasoning and reflective

---

[61] See Laurie A. Paul, *Transformative Experience* (Oxford, UK: Oxford University Press, 2014).

endorsements that *should* be available. And the emotional conditions orchestrated by ISIS indoctrinators plausibly impose a similar reasons-independent constraint on reflection and reflective endorsement. But, in these indoctrination cases, a conversion to a new way of life ends up being produced because the person is, paradoxically, drawn by critical scrutiny and driven by the search for truth, on matters that they are blocked from scrutinizing in salient alternative ways.

Thus, I find it very plausible that the reasoning of the youngsters I am concerned with is unduly constrained by emotion, because these individuals become blocked by overwhelming emotions that they cannot handle, from consciously reasoning in certain ways, and from reflectively endorsing beliefs, pro-attitudes, and practical commitments that clash with jihadism.

Consider again Andrea Westlund's condition of answerability to external critical perspectives. Perhaps we should think that the youngsters I am concerned with, during the early stages of their radicalization process, are *too* responsive (in both their theoretical and practical reasoning) to the justificatory demands of *certain* external critical perspectives. It is not that they are too ready to defer to certain people (or books, or institutions) with those perspectives; this happens, if ever, only late in the process. Rather, the external perspectives with which ISIS indoctrinators constantly confront these youngsters are perceived by them as urgently in need of justificatory responses that they fear they cannot fully provide.

In any case, whatever the principled explanation is for why this indoctrination undermines autonomy, it seems very plausible that it does undermine autonomy, and that it does so without undermining any of the rational, emotional, and social conditions that have already been identified by different historical accounts in the literature (on a broad reading of "historical accounts," on which classical and relational accounts also count as historical). Thus, if one holds a historical account of autonomy, one should also accept an emotional condition on doxastic and volitional autonomy that rules out the emotional states and conditions that are orchestrated by this kind of indoctrination when it is effective. It would be best to find a principled reason to rule out these kinds of emotional states and conditions, but, even lacking such a principled reason, we should at least add to our accepted list of historical conditions on autonomy, the absence of these emotional constraints on critical reflection, practical deliberation, and the considered endorsement of views, values, desires, goals, and action-guiding commitments.

VI. Conclusion

I argued that recent fieldwork by psychologists and anthropologists on the kind of indoctrination that ISIS uses to recruit Westerners justifies introducing a necessary condition on autonomy that is not implied by the

conditions postulated by existing accounts: For a person to remain autonomous regarding a given subset of her views, values, goals, practical commitments, and so on, her endorsement of these must not be constrained by a constant state of fear, anxiety, guilt, paranoia, and the like, and by a hypertrophied perception of the duty and urgency of offering justifications of one's views and pro-attitudes in answer to external justificatory challenges that provoke these distressing emotions.

More empirical work is needed to continue learning about the role that emotions play on severe indoctrination at different stages of cognitive and emotional development. The better we understand how these influences work, the better we will be able to prevent, stop, and reverse them. Despite the wide diversity of paths to terrorism, are there pre-existing conditions, psychological or other, that make some people more prone to being effectively indoctrinated into violent extremism? For instance, do people who get indoctrinated have a tendency to be too attached or irrationally attached to their views, in such a way that they often disregard contrary evidence and give too much weight to confirmatory evidence? Are people who tend to believe in conspiracy theories less rational than other people (less inclined to conscious reflection and so on), or are they rather *hyper-rational*, namely more prone to seek unificatory explanations of disparate phenomena, in such a way that no single event remains outside the overarching theory in question? These are important empirical questions that still need to be answered.

Even though I wrote this essay without the answers to these questions, I believe that the empirical information we have so far is enough to justify the conclusion that there are distinctively emotional conditions on autonomy that have not been identified in the literature. The emotional conditions that are postulated and implied by existing accounts are not enough to yield (together with rational and social requirements) sufficient conditions for autonomy.

Philosophy, Universidad de Los Andes