

Epistemic Paternalism, Personal Sovereignty, and One's Own Good

Michel Croce (University of Edinburgh & University of Lisbon)

Forthcoming in G. Axtell, A. Bernal (eds), *Epistemic Paternalism Reconsidered: Conceptions, Justifications, and Implications*, Rowman & Littlefield, 2020.

Abstract. A recent paper by Bullock (2016) raises a dilemma for proponents of epistemic paternalism. If epistemic paternalists contend that epistemic improvements contribute to one's wellbeing, then their view conflates with general paternalism. Instead, if they appeal to the notion of a distinctive epistemic value, their view is unjustified, in that concerns about epistemic value fail to outweigh concerns about personal sovereignty. In this chapter, I address Bullock's challenge in a way that safeguards the legitimacy of epistemic paternalism, albeit restricting its scope to a limited range of cognitive projects. After shedding light on a problem with how Bullock singles out cases to which the dilemma applies, I argue that there is at least one reasonable way of interpreting the notion of 'personal autonomy' which legitimates and justifies undertaking epistemically paternalistic interferences for one's epistemic good.

1.

In recent years, philosophical discussions about the nature and the legitimacy of paternalistic interferences have ventured into the epistemological domain, where questions have been raised in relation to a cognate practice, namely *epistemic paternalism*. Epistemic paternalism is the thesis according to which in some circumstances, we are epistemically justified in interfering with the inquiry of others for their own epistemic good without consulting them on the issue (Ahlstrom-Vij 2013: 4).

The bulk of the recent discussion on epistemic paternalism has concerned when and how interferences of this sort are legitimate or, to put it differently, when someone can be justified in undertaking them. Despite a widespread worry that epistemic paternalism is bound to be a practice that undermines our epistemic autonomy—if not our freedom, in general—several epistemologists have attempted to offer an account of legitimate or justified epistemic paternalism (Ahlstrom-Vij 2013; Bullock 2018; Goldman 1991; Pritchard 2013) or have inquired into who is rationally entitled to undertake epistemically paternalistic interferences, and in virtue of which features one inherits this entitlement (Croce 2018).

Somewhat less attention has been directed toward the issue of how to motivate epistemic paternalism, that is, of providing an explanation of *why* epistemically paternalistic interferences are justified. Addressing this question involves giving an account of the value of the epistemic improvements that these interventions are meant to bring about. In her “Knowing and Not-knowing for Your Own Good: The Limits of Epistemic Paternalism”, Emma Bullock raised a neat dilemma for epistemic paternalists, namely one that puts pressure on the very possibility of offering a motivation for paternalistic interferences in the domain of inquiry. In a nutshell, if epistemic paternalists contend that epistemic improvements contribute to one’s wellbeing, then their view conflates with general paternalism. Instead, if they appeal to the notion of a distinctive epistemic value, their view is unjustified, in that concerns about epistemic value fail to outweigh concerns about personal sovereignty.

The goal of this chapter is to address Bullock’s challenge in a way that safeguards the legitimacy of epistemic paternalism, albeit restricting its scope to a limited range of cognitive projects. After briefly recalling the definition of epistemic paternalism and reconstructing Bullock’s dilemma (§2), I shall attempt to fulfill the chapter’s aim by way of two main moves. First, I shall take issue with the setup of Bullock’s dilemma and, in particular, with how Bullock singles out cases to which the dilemma applies. Although I agree with her that the dilemma should not apply to standard educational settings involving young children—since in most cases of this sort, epistemic paternalism is justified—I shall point out that her proposed criterion for excluding these cases is highly problematic (§3). Then, I shall offer a solution to the dilemma on behalf of those who find scope for justification of epistemic paternalism in the distinctive value of some interferences. Specifically, I shall argue that there is at least one reasonable way of interpreting the notion of ‘personal autonomy’ which legitimates and justifies undertaking epistemically paternalistic interferences for one’s epistemic good (§4).

According to Ahlstrom-Vij's (2013) recent definition, a paternalistic interferer (henceforth, PI) undertakes an epistemically paternalistic practice towards a subject interfered with (henceforth, S) by doing (or omitting to do) X if and only if the following conditions are met:

- (i) Doing X interferes with the epistemic autonomy or freedom of S to conduct inquiry in whatever way S sees fit (*interference condition*);
- (ii) PI does so without consulting S on whether S should be interfered with in the relevant manner (*non-consultation condition*);
- (iii) P does so for the purpose of making S epistemically better off (*improvement condition*).

Let us briefly comment on these requirements.¹ The interference condition individuates the nature of the phenomenon we are dealing with, that is, an intervention with the way in which someone else decides to manage their epistemic life. The non-consultation condition illustrates that for an interference to be epistemically paternalistic, PI does not ask S whether S welcomes PI's interference. Specifically, PI's interference is paternalistic irrespective of whether S would object to it, had S been consulted on the issue. The improvement condition sets the bar for the legitimacy of epistemic paternalism, as it specifies that the interference has to be aimed at bringing about an epistemic benefit for S—where this epistemic benefit can be cashed out along veritistic lines (Goldman 1991; Ahlstrom-Vij 2013) or along a broader perspective that encompasses S's growth in understanding or in intellectual virtues (Croce 2018; Pritchard 2013).

The main difficulty with any attempt to motivate epistemic paternalism concerns the general account of epistemic value from which paternalistic interferences inherit their justification and its relationship with the broader notion of one's own good or wellbeing. Bullock distinguishes two approaches to epistemic paternalism depending on how each determines what counts as a genuine benefit to the subject interfered with. *Eudaimonic* epistemic paternalism confers value to paternalistic interferences based on their contribution to S's *overall wellbeing*. *Strict* epistemic paternalism, in contrast, confers value to paternalistic interferences based on the *epistemic benefit* they generate for S.

The two horns of Bullock's dilemma are basically arguments for why both views fail to provide a compelling motivation for undertaking paternalistic interferences in the epistemic domain.

Before introducing her arguments, it should be pointed out that Bullock's main targets are cases in which the individual interfered with is the one who benefits from the intervention—i.e., cases of *direct* epistemic paternalism—and is mature enough to make decisions about their own inquiry—i.e., cases of *hard* epistemic paternalism (2018: 434-435).²

Let us first consider Bullock's argument against eudaimonic epistemic paternalism. On this view, paternalistic interferences can generate epistemic improvements that are constitutive of wellbeing or merely instrumentally valuable to wellbeing if they simply increase the likelihood that S will be better off in terms of their wellbeing (Bullock 2018: 437). Either way, Bullock argues, eudaimonic epistemic paternalists will have a hard time showing how their view distinguishes itself from general paternalism. For not only cases of general paternalism fulfill the interference condition and the non-consultation condition. If we read the improvement condition along the lines of eudaimonic epistemic paternalism, some general paternalistic interventions fulfill it too, insofar as generating epistemic improvements is just one of the ways in which an interference can improve S's wellbeing.

But why should epistemic paternalists worry about keeping a neat distinction between the eudaimonic account and general paternalism? As Bullock rightly notes, if epistemic paternalism collapses into general paternalism, it is no longer clear what role the improvement condition is supposed to play (438). For it may well be the case that some paternalistic interventions improve S's overall wellbeing at the cost of sacrificing S's epistemic welfare and that granting legitimacy to such interferences undermines the scope of epistemic paternalism.

To name just one of Bullock's examples, self-enhancement bias can contribute to one's wellbeing by favoring an overly positive conception of oneself and thereby diminishing the risks of depression (439). Eudaimonic epistemic paternalists seem committed to granting that there may be cases in which interferences that promote this cognitive bias are justified in virtue of their

contribution to S's overall wellbeing, despite the fact that such bias reduces or obstructs S's self-knowledge.

Strict epistemic paternalism is well positioned to avoid this unwelcome consequence of the eudaimonic account, in that it assumes that epistemically paternalistic interferences fulfill the improvement condition to the extent that they bring about a genuine epistemic benefit for S. Thus, strict epistemic paternalists—at least, on the veritistic approach Bullock seems more concerned with—would not concede that promoting self-enhancement bias can constitute a legitimate paternalistic intervention in the epistemic domain. Nonetheless, strict epistemic paternalism has to face a challenge of its own, in that it has to explain whether and how epistemically paternalistic interferences can bring about epistemic benefits without damaging S's overall wellbeing. Bullock's contention is that strict epistemic paternalists fail to provide a compelling argument of this sort.

Available answers to the challenge for strict epistemic paternalism go in two different directions. One is offered by the *balancing goods condition*, according to which an epistemically paternalistic interference is justified insofar as PI has no good reason to suppose that their intervention will make S all-things-considered worse off. In other words, this condition requires that PI weighs the epistemic benefits of their intervention with its other possible non-epistemic effects and confers justification to it to the extent that the reasons for intervention outweigh the reasons against interfering.

This requirement provides two important advantages, in that it accommodates cases in which PI's interference aims at making S better off both epistemically and in some other relevant respect, and it avoids the risk that PI's interference be justified when it brings about an epistemic benefit for S at the cost of damaging other spheres of S's wellbeing.³ However, as Ahlstrom-Vij points out, in several cases weighing reasons might become an extremely difficult task because it might not be at all obvious how to put on the same scale different *quantities* of values that differ in *quality* too (2013: 116-117).

To avoid this problem, Ahlstrom-Vij suggests to replace the balancing goods condition with the *alignment condition*, according to which an epistemically paternalistic interference is justified only insofar as PI's epistemic reasons for the interference are aligned with PI's non-epistemic reasons for the interference, either by being additional reasons for intervening or by being neutral—that is, by not constituting reasons against intervening (117). Unlike the former condition, the alignment condition merely requires that PI only knows the valence of the reasons involved in the evaluation—namely, the direction for or against a given interference.

Yet, this condition is problematic too, as there might be cases in which a weak non-epistemic reason against intervening would commit us to deny that an interference in favor of which we have strong epistemic and non-epistemic reasons can be justified. For example, suppose the surgical procedure through which I can donate my bone marrow to a friend is potentially dangerous given my actual medical conditions, and I tell the doctor I want to do it and do not want to be informed about any immediate or long term side effect I could suffer from. The doctor initially consents to my request, but after acknowledging that the surgery went just fine and reflecting that my decision not to know will sit badly with my hypochondriac tendencies, she disregards my request and informs me about the positive results of the procedure.

In this case, the doctor has a weak non-epistemic reason against intervening, i.e. my request not to be informed, and strong epistemic and non-epistemic reasons in favor of intervening, both having to do with the positive effects of knowing how the surgery went. It seems reasonable to suppose that the doctor's epistemically paternalistic interference is justified as the benefits of intervening override the harm of disregarding my request. If so, then the alignment condition fails to provide a necessary condition on justified epistemic paternalism. Ahlstrom-Vij is ready to concede that there might be a very minor range of situations to which the alignment condition does not apply. However, Bullock has shown that this weakness is more pervasive than Ahlstrom-Vij thinks. Basically, no paternalistic interference would be justified on this view because all interventions of this kind would have to outweigh one clear reason against interfering, namely the violation of S's

personal sovereignty. For it should be clear that limiting or disrespecting one's personal sovereignty is a form of damaging one's wellbeing (2018: 441-442).

According to Bullock, this problem undermines the alignment condition, for neither of the replies available to the strict epistemic paternalist meets the challenge. On the one hand, strict epistemic paternalists could deny that we have to respect personal sovereignty, but this is an extreme view that would need to be supported by a strong argument. On the other, they could contend that personal sovereignty is only *pro tanto* valuable but, in that case, only weighing reasons would allow us to determine whether in a given situation the concern with personal sovereignty is outweighed by other available reasons for interfering with one's inquiry.

Since the alignment condition cannot be rescued from its weaknesses, it looks as though strict epistemic paternalists will have to address the concern with personal sovereignty with the resources provided by the original balancing goods condition. In this scenario, Bullock argues that they could either take an *extreme* route and contend that bringing about an epistemic benefit for S is the only good that can outweigh personal sovereignty or opt for a *moderate* position. In particular, strict epistemic paternalists could argue that (a) the concern with personal sovereignty can be addressed by interventions that promote S's epistemic benefit *or* S's wellbeing, or (b) the concern with personal sovereignty can be addressed by interventions that promote *both* S's epistemic benefit *and* S's wellbeing (2018: 442).

All these views have their own problems. On the extreme position, it is far from clear why epistemic value should be the only kind of value that can outweigh personal sovereignty. Proponents of moderate option (a) shall show how the promotion of an epistemic benefit can trump personal sovereignty and the promotion of one's wellbeing when this contrasts with epistemic value. Finally, proponents of moderate option (b) shall show why it is necessary that an interference brings about both epistemic and non-epistemic value in order for epistemic paternalism to be justified.

The challenges for proponents of the moderate options, as Bullock admits, can in principle be met and therefore constitute the only available opportunity for epistemic paternalists—in fact, for

strict epistemic paternalists—to solve the dilemma and justify epistemic paternalism as a legitimate practice (444). The remainder of this chapter will be devoted to outlining an argument in favor of this position.

3.

Before addressing the dilemma and outlining a potential way out for strict epistemic paternalists, I want to shed light on a general, yet related, issue that seems to be receiving less attention than it deserves in the current discussion on epistemic paternalism. As I pointed out in the last section, Bullock's dilemma takes stock of hard epistemic paternalism, that is, interventions aimed at bringing about some sort of epistemic benefit for a subject who is judged "to be worthy of having their decisions about their inquiry respected" (2018: 436). Bullock intentionally sets aside cases of soft epistemic paternalism, that is, circumstances involving individuals who lack the competence to make decisions about their own inquiry, and offers the case of a child's education as a paradigmatic example for this type of paternalistic interventions.

As I shall argue below, in principle there is nothing wrong with this restriction in scope. For as we have already made clear, this phenomenon is in itself a violation of one's autonomy and therefore a *prima facie* problematic intervention. Thus, we should hope that the epistemically paternalistic interferences that take place in a given environment are meant to bring about an epistemic benefit only for individuals who lack the ability to make decisions about their own inquiry.

However, a closer look at the requirements of soft epistemic paternalism reveals that Bullock's criterion for distinguishing between hard and soft instances of epistemic paternalism looks more problematic than it initially appeared. The main problem with this criterion lies in the very notion of a competence to make decisions about one's own inquiry. Bullock takes education as a paradigmatic domain to clarify the distinction between the two kinds of epistemic paternalism, but education does not seem to fit the criterion. If it is clear that the child lacks the competence to make decisions about their own inquiry, it is not clear what it takes for a mature epistemic subject to have such competence.

Consider a case in which a parent moves a joke history book out of their child's reach to prevent them from acquiring inaccurate historical information (Pritchard 2013: 15), or a case in which a teacher refrains from providing a student with the answer to a geometry problem in order to let them work out a solution and develop analytical skills. It seems plausible to contend that both situations constitute cases of soft epistemic paternalism, in that the subjects interfered with lack the competence to make decisions about their own inquiry and the educators' paternalistic interventions bring about an epistemic benefit for them.

Now consider another pair of cases involving mature epistemic subjects. In the first scenario, a brilliant math student based at MIT is approaching for the first time Hilbert's third problem—i.e., the first of Hilbert's problems to be solved within a year of its formulation. The professor does not reveal to the student that this is one of Hilbert's solved problems, nor does she provide him with the solution because she judges that letting the student work on the problem on his own would allow him to develop a better understanding of the problem and improve their mathematical abilities. In the second scenario, two colleagues at a big company are in charge of recruiting twenty new employees and are evaluating a huge number of CVs to shortlist forty applicants. Kate knows that Jerry has developed a bias against female applicants and therefore decides to provide him with blind CVs in order to prevent his bias from affecting his judgment. Should we consider these examples as cases of hard epistemic paternalism? That is: are the student and Jerry mature epistemic subjects who have the competence to make decisions about their own inquiry?

At first glance, it looks as though we should reply in the affirmative: the subjects interfered with are neither children nor individuals with limited awareness of their autonomy in both the practical and the intellectual sphere. However, their competence to make decisions in the respective domains is severely limited. Given the early stage of his career, presumably the math student does not know whether it would be epistemically better for him to get a straightforward solution to Hilbert's problem or to approach it as an extant problem in need of a solution. For he lacks the competence

to evaluate what he could learn by dealing with the problem, what skills he would develop, and how this effort would reward him later in his career.

Similarly—in fact, even more dramatically—Jerry is completely blind to his cognitive bias. In this case, we should be wary of granting him the competence to decide about his own inquiry for two main reasons. First, because he lacks awareness of something relevant to possessing such competence. Second, because—as in the nature of biases in general—he would be reluctant to admit that he has developed one and/or is unable to change his epistemic conduct, were Kate to make him aware of his attitude towards female applicants.

Yet, if the characters of our toy examples lack the ability to choose what is epistemically better for them to do in the respective contexts, then it becomes at least dubious whether we should treat these cases differently than the ones involving children. As this argument illustrates, the boundaries between soft and hard epistemic paternalism become less and less neat. Hence, we are entitled to wonder what consequences this has on the kinds of epistemically paternalistic interferences that should be immune to Bullock's dilemma.

A tempting answer would be to extend the domain of soft epistemic paternalism so as to encompass many educational cases involving mature-yet-incompetent epistemic subjects. This move should not worry us too much, in that the overarching aim of education is no doubt epistemic and its main implications are typically confined to the intellectual domain.⁴ The bad news for Bullock is that, as the cognitive bias example shows, the problem of separating soft and hard cases of epistemic paternalism can be pressing even outside education, where one's decisions may also affect the lives of others.

Thus, on the one hand, it seems as though there is scope for restricting the set of interferences that fall within soft epistemic paternalism as much as possible. On the other hand, Bullock's criterion for distinguishing between soft and hard cases of epistemic paternalism authorizes us to include a larger range of interferences in the set of soft cases, thereby exempting them from falling prey to Bullock's own dilemma. I must confess that I have no ready-made answer to this problem,

which, to my knowledge, has been surprisingly overlooked to date. This problem, though, becomes somewhat relevant in relation to my proposed solution to Bullock's dilemma. So, I shall postpone any further comment on it until the concluding paragraphs of the next section and focus my attention on the discussion of Bullock's dilemma.

4.

The solution I shall propose to escape the dilemma raised by Bullock constitutes an argument on behalf of strict epistemic paternalists embracing moderate option (a), as discussed in §2. More precisely, I will argue that in some cases epistemic paternalism can be justified because the promotion of an epistemic benefit trumps the concern with personal sovereignty both when such epistemic benefit aligns with one's overall wellbeing and when it clashes with one's wellbeing.

Let me be clear from the start: I am not going to sell the balancing goods condition as an easy requirement to fulfill here, as I have already clarified that it is far from being so. My goal is rather to single out a set of cases of (hard) epistemic paternalism whose structure is such that they do not fall prey to Bullock's dilemma if one is willing to endorse a broader account of personal autonomy than mere personal sovereignty. This result might look modest to a demanding reader who seeks a wide spectrum treatment for the issue at stake. Yet, it will provide us with enough concrete instances of justified epistemic paternalism to manage and, as I argued in §3, we might well have good reasons to keep this list relatively short.

The *personal sovereignty model* of personal autonomy revolves around "the idea of having a domain or territory in which the self is sovereign" (Feinberg 1976: 52). The boundaries of this territory presumably include any self-regarding decision, that is, any decision that directly affects only the interest of the decision maker (56). Any violation of the boundaries by uninvited individuals is illegitimate. If we take this account by the book, the right of self-determination is as morally basic as the good of self-fulfillment. What is more, self-determination may lead one to harm oneself and yet nobody else has a right to interfere with one's decisions, in that 'autonomy is even more important than personal wellbeing' (59).

Clearly, this model of personal autonomy allows no room for any sort of paternalistic interference, be it general or epistemic. It might well be the case that Jerry's epistemic agency, in our example, would benefit from having personal information about the applicants removed from their CVs, but his right of self-determination outweighs Kate's concerns with his epistemic welfare, despite the fact that it will lead him to provide a biased assessment of the CVs.⁵ Thus, if this is how we conceive of personal autonomy, (epistemic) paternalism will never be justified.

However, other accounts of personal autonomy provide us with a different diagnosis of the relationship between self-determination and personal good, namely one in which personal sovereignty is only *pro tanto* valuable and, as a consequence, epistemically paternalistic interferences may be justified on specific grounds. Let us consider the *model of autonomy as a condition* outlined by Feinberg (1989), according to which autonomy is a *condition* that involves both the ability to make rational choices and *de facto* self-government, that is, the opportunity to exercise one's rights and capacities (31). According to this approach, autonomy requires several virtues: among them, Feinberg mentions moral authenticity, moral independence, integrity, self-discipline, self-reliance, but also those intellectual virtues that make one responsible for the self, namely intellectual courage, trustworthiness, reliability, and good judgment (44).

If we assess Jerry's example in light of this account, it is no longer clear that Jerry is such an autonomous individual. His overwhelming proneness to cognitive biases might undermine his ability to make rational choices and it certainly prevents him from displaying relevant virtues such as reliability, good judgment, and self-discipline—to the extent that being prone to his cognitive biases makes him 'governed from the outside' (40).

However, if it is true that Kate's interference brings about an epistemic benefit for Jerry, our critic might still object that such benefit fails to counterbalance the violation of Jerry's right to self-determination (or personal sovereignty), which is necessary to provide strict epistemic paternalists with a way out of Bullock's dilemma. While this diagnosis does not suffice to settle the problem on its own, it points us in the right direction: as I shall argue in the remainder of this section, epistemic

benefits trump concerns about personal sovereignty insofar as they contribute to one's overall personal autonomy, no matter whether on balance they also improve one's overall wellbeing,

Two brief remarks are in order here: first, following the model of autonomy as a condition, I assume that having autonomy is a matter of degree; second, I shall contend that becoming (more) autonomous may sometimes happen at the expense of one's wellbeing. Let me clarify these points while discussing another example involving a specific kind of social epistemic structure: *epistemic bubbles*.⁶ Epistemic bubbles are structures of exclusion that prevent large groups of epistemic subjects from taking into due account some kinds of information. They impede distribution of a complete range of information by omitting relevant testimony from sources endorsing a rival perspective and therefore they distort the informational environment of their members. Bursting epistemic bubbles requires that some members are exposed to excluded information.

Epistemic bubbles are particularly interesting in the discussion of epistemic paternalism because they not only degrade the epistemic welfare of their members, but they are also likely to put members' personal autonomy at risk. Consider the following example:

THE CIA BUBBLE. Suppose Sarah has been raised in a small rural community that believes in all sorts of conspiracy theories regarding the corruption of the US government and the involvement of the CIA in events such as 9/11, the Kennedy assassination, and Malala Yousafzai's attempted murder. Sarah has been convinced that these conspiracy theories are reliable by made-up data and evidence as well as by acknowledging that all members of her community endorse these theories. After living several years in a small village, Sarah moves to the city to work and befriends Mary, who does not seem to have a problem avoiding any conversation about the U.S. government and the CIA, as Sarah explicitly requests. One day, Mary overhears a phone call between Sarah and her family and figures out that Sarah belongs to a community with the aforementioned features. Mary decides to intervene by inviting Sarah for dinner and presenting her with pieces of counterevidence for her

conspiracy theories as well as with journal articles detailing the mechanisms typical of epistemic bubbles and their likelihood to proliferate in communities with such and such features.

Sarah's community constitutes an example of an epistemic bubble, in that their members have not been exposed to any sort of counterevidence and foster their belief in these conspiracy theories by discussing these issues solely with insiders. Furthermore, no matter the outcome of Mary's interference, this story exemplifies a case of epistemic paternalism: Mary infringes Sarah's personal sovereignty and freedom to conduct her inquiries as she wishes to burst the bubble and improve her epistemic welfare. Is her interference justified?

On the personal sovereignty model of autonomy, Mary's intervention is clearly unjustified, as Sarah's right of self-determination is more valuable than her epistemic welfare and Mary's interference has infringed this inviolable right. Nonetheless, I believe that this diagnosis of the case is a complete loss. Intuitively, the epistemic gain of being freed from a deeply rooted detrimental condition—such as that of being trapped in an epistemic bubble—outweighs the harm of violating of one's right of self-determination. Moreover, Mary's intervention positively contributes to improving Sarah's personal autonomy, when conceived along the lines of the model of autonomy as a condition.

For it might well be the case that Sarah displays all the moral virtues that contribute to one's possession of autonomy, but she no doubt fails to display the required intellectual virtues. By omitting their members' exposure to all the available information and counterevidence, epistemic bubbles compromise the members' reliability and good judgment. By favoring interaction among like-minded individuals, epistemic bubbles bootstrap corroboration of the relevant theories or information, thereby deluding their members into believing that they are free to direct their own lives when, in fact, they are only exposed to a limited range of possibilities.

Thus, by bursting the epistemic bubble, Mary not only brings about an epistemic benefit for Sarah, but she also puts Sarah in a position to improve her personal autonomy. Sarah can now (i)

acknowledge the functioning of epistemic bubbles and their detrimental effects on one's agency, (ii) inquire into the responsibility of those (if any) people who put her in the bubble and those who actively sustain this structure, and (iii) develop those virtues that are required to become an autonomous person. For these reasons, we can conclude that Mary's paternalistic interference is justified.⁷

Someone might rightly point out that, despite addressing the issue of whether epistemic value can trump one's personal sovereignty, my argument does not solve Bullock's dilemma, in that it remains silent about whether epistemic value can counterbalance considerations of wellbeing. As a matter of fact, nothing in *THE CIA BUBBLE* elicits the conclusion that this is a case in which Mary's epistemic reasons for intervening clash with the non-epistemic concern with Sarah's overall wellbeing.

As I anticipated early on in the section with the second remark, I concede that becoming more autonomous may negatively affect one's overall wellbeing. For example, it might be the case that, for an individual like Sarah, the non-epistemic value of becoming more autonomous fails to outweigh the psychological costs of getting out of such a longstanding bubble—e.g., the affective costs of questioning the value of important personal relationships, combined with the distress and anxiety that this process generates. Sarah's life is just a mess now that she has realized that the community in which she has been raised was obstructing her access to knowledge and was instilling all this bullshit about the government's corruption and the CIA's bloody activities! No doubt she has grown in autonomy but, in the end, her discoveries did not change her life for the better, while the psychological effects of such discoveries changed her life for the worse.

This case exemplifies all the relevant features of Bullock's challenge for strict epistemic paternalists endorsing moderate option (a). For the epistemically paternalistic interference brings about an epistemic benefit and improves one's personal autonomy at the cost of violating one's personal sovereignty and conflicting with one's overall wellbeing. Would such an interference be justified?

I would answer this question in the affirmative, for this reason: to the extent that an epistemically paternalistic interference contributes to fostering one's autonomy, it helps the subject interfered with to reach a position at which they can freely choose how to manage their own wellbeing. Thus, although the epistemic reasons in favor of interfering do not align with the non-epistemic considerations of one's wellbeing, they counterbalance this concern by freeing one from any external conditioning on how one should manage their own epistemic and non-epistemic life.

This way out of Bullock's dilemma has two important features: first, it allows strict epistemic paternalists to stick to the idea that epistemic value can trump the concern with one's personal sovereignty while taking into consideration the value of personal autonomy; second, it restricts the range of justified epistemically paternalistic interferences to those interventions that promote the autonomy of the subject interfered with. Thus, on this view, playing physics lectures to a sleeping individual to improve their epistemic welfare will not be justified because the interference, despite allegedly bringing about some epistemic benefit for S, does not contribute to S's overall personal autonomy (see Bullock 2018: 442-443).

As a conclusion to this contribution, I shall raise a final objection to the proposed solution to Bullock's dilemma. If PI's epistemic interference is justified insofar as it improves S's personal autonomy, so the objection goes, then it is not clear whether S had the competence to make decisions about S's own inquiry in the first place. But this undermines the argument I have proposed on behalf of strict epistemic paternalists because it means that the cases considered herein are instances of soft—rather than hard—epistemic paternalism.

We are finally back to the problem I raised in §3 about the boundaries of the distinction between soft and hard cases of epistemic paternalism. There are two available conclusions to our journey. On one conclusion, I have offered an argument that grants justification to a restricted range of cases of hard epistemic paternalism, namely those that, besides improving one's epistemic welfare, foster one's personal autonomy. On the other, I have failed to do so because the cases of epistemic paternalism with which I was concerned are, by Bullock's own standards, cases of soft epistemic

paternalism and therefore are already immune to Bullock's dilemma. Either way, the end of the story is a good one: for, besides standard cases in the educational domain, there is a relevant set of scenarios in which our due concerns with promoting one's epistemic welfare can justify our interventions despite the violation of one's personal sovereignty we inevitably incur.

NOTES

¹ For a more detailed analysis of these conditions see, e.g., Bullock (2018: §1) and Croce (2018: §2).

² Bullock contrasts these notions with those of *indirect* epistemic paternalism, i.e., interferences with an individual that aim to benefit others, and *soft* epistemic paternalism, i.e., interferences with an individual who lacks the competence to make decisions about their own inquiry.

³ Ahlstrom-Vij calls this form of interferences *mixed paternalism* (2013: 115). An example of mixed paternalism is a judge's decision not to reveal information about a defendant's past record of crimes both to prevent the jurors from developing a bias against the defendant (epistemic benefit) and to safeguard the welfare of the defendant themselves (non-epistemic benefit).

⁴ It remains a highly disputed question in the epistemology of education what constitutes the primary epistemic aim of education (see, e.g., Baehr 2019; Kotzee, Carter, Siegel 2019; Siegel 2018). What matters for the purposes of this chapter is simply to note that, despite ongoing divergences, epistemologists of education agree that the primary or overarching aim of education is an epistemic one.

⁵ Notice, though, that Jerry's right of self-determination does not outweigh the candidates' right to receive a fair assessment of their suitability for the job. Thus, although it might not be legitimate for Kate to interfere with Jerry's agency for his own epistemic good, it might well be legitimate for her to intervene in order to protect the candidates' rights.

⁶ The following characterization of epistemic bubbles is taken from Nguyen (2018).

⁷ It is worth pointing out that a similar argument can be run for interferences with the agency of members of another detrimental epistemic structure, namely *echo chambers* (Nguyen 2018).

BIBLIOGRAPHY

- Ahlstrom-Vij, Kristoffer. 2013. *Epistemic Paternalism: A Defence*. Basingstoke: Palgrave Macmillan.
- Baehr, Jason. 2019. "Intellectual Virtues, Critical Thinking, and the Aims of Education." In M. Fricker, P. Graham, D. Henderson, N. Pedersen & J. Wyatt (eds.), *The Routledge Handbook of Social Epistemology*. London: Routledge, pp. 104-121.
- Bullock, Emma. 2018. "Knowing and Not-Knowing For Your Own Good: The Limits of Epistemic Paternalism." *Journal of Applied Philosophy* 35(2): 433-447.
- Croce, Michel. 2018. "Epistemic Paternalism and The Service Conception of Epistemic Authority." *Metaphilosophy* 49(3): 305-327.
- Feinberg, Joel. 1989. *The Moral Limits of The Criminal Law Volume 3: Harm to Self*. Oxford: Oxford University Press.
- Goldman, Alvin. 1991. "Epistemic Paternalism: Communication Control in Law and Society." *The Journal of Philosophy* 88, no. 3:113-131.
- Kotzee, Ben, Carter, Adam, and Siegel, Harvey. 2019. "Educating for Intellectual Virtue: A Critique from Action Guidance." *Episteme*. doi: [10.1017/epi.2019.10](https://doi.org/10.1017/epi.2019.10)
- Nguyen, Thy. 2018. "Echo Chambers and Epistemic Bubbles." *Episteme*. doi: [10.1017/epi.2018.32](https://doi.org/10.1017/epi.2018.32)

Pritchard, Duncan. 2013. "Epistemic Paternalism and Epistemic Value." *Philosophical Inquiries* 1, no. 2:9-37.

Siegel, H. (2018). *Education's Epistemology: Rationality, Diversity, and Critical Thinking*. New York: Oxford University Press.