

An Epistemological Account of Indoctrination

Elena Silvia Ludovica Raimondi

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Philosophical Studies MPhil Stud
of
University College London

Department of Philosophy
University College London

2021

Statement of originality

I, Elena Silvia Ludovica Raimondi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

What is indoctrination? This work aims to persuade the reader that a promising approach to answering this question is conducting—in Haslanger’s (2000) terminology—a descriptive analysis of indoctrination from an epistemological viewpoint. To do so, I argue that the epistemological viewpoint is a privileged one because it encompasses all instances of indoctrination, and that a descriptive kind of analysis is the most appropriate to improve our understanding of an often mischaracterised phenomenon. Further, to show that the suggested approach is fruitful, I develop an account of indoctrination along these lines.

I develop my account in dialogue with the literature on indoctrination in Analytic Philosophy, where indoctrination is generally understood as a degenerate form of proper education, and where efforts are mostly devoted to explicating why indoctrination deviates from proper education. Against this tradition, and borrowing from the literature on manipulation instead, I argue that indoctrination is rather the exploitation of an epistemic-cum-cognitive vulnerability. It consists not in inculcating beliefs, but in instilling a defective pattern of response to reasons, which leads the subject of indoctrination to systematically choose to do the manipulator’s bidding. The challenge then is not distinguishing indoctrination from proper education, but from other defective forms of reasoning.

Impact statement

This thesis is meant to be a step in the direction of better understanding the phenomenon of indoctrination—what it is, how it works, and what harms it causes. Alongside contributing to the academic debate on the topic, it aims to offer to society at large a more technical account of indoctrination than its ominous impressions in popular culture, and one that is potentially more useful in counteracting indoctrination effectively and respectfully.

Acknowledgements

I would like to thank my supervisor, Dr Robert Simpson, for his indispensable guidance in the writing of this thesis. I am deeply grateful for your mentorship and for everything you taught me about writing in Philosophy. It has been a pleasure to work with someone so reliable and inspirational.

Many thanks to Niels Boecker, who contributed to the making of this thesis in various ways, including showing me techniques to track my progress, helping me with the technological side of things and motivating me to stay focused.

I am ever thankful to my parents and brother for unconditionally supporting me throughout any project I decide to embark on. Thanks to my aunt for sparking my interest in the topic of indoctrination years ago by gifting me a short novel.

Finally, I owe my gratitude to the Collegio Nuovo of Pavia for enabling me to gain experience that opened doors for me, and Dr Ian Carter of the University of Pavia, who supervised my Bachelor's thesis and encouraged me to apply to the Philosophical Studies MPhil Stud programme at UCL in the first place.

Contents

Introduction	13
1 Proposing an approach	15
1.1 Why an epistemological account?	15
1.2 A descriptive analysis	20
2 Drawing from the literature	25
2.1 Indoctrination in the literature	25
2.2 Beyond the tripartite model	35
2.3 Yaffe’s account	41
3 Developing the account	49
3.1 The emergence of a defective pattern	49
3.2 One pattern, different power structures	58
3.3 A delicate balance	68
3.4 Revising our worries about indoctrination	73
3.5 Exploiting a vulnerability of the mind	75

4 Ideas for counteracting indoctrination	79
4.1 Dislodging the defective pattern	79
4.2 Belief polarisation	81
4.3 The polarising effect of indoctrination	85
4.4 A further harm of indoctrination	88
4.5 Two strategies to bypass the defective pattern	92
Concluding remarks	99
Bibliography	108

Introduction

What is indoctrination? The purpose of this work is to persuade the reader that a promising approach to answer this question is to focus on the epistemological aspect of indoctrination by conducting a descriptive analysis. That is, in a nutshell, to look into the paradigm cases of indoctrination, and capture what goes on from an epistemological point of view. Further, to show that this approach is fruitful, I will develop an account of indoctrination along these lines and argue that it offers new insights into the phenomenon.

This is how I will proceed: In chapter 1, I will discuss the advantages of assuming an epistemological viewpoint to account for indoctrination, and I will explain what a descriptive analysis consists in by referring to Haslanger's (2000) framework to understand what counts as a philosophical analysis of a concept.

In chapter 2, I will dig into the literature on indoctrination in the Analytic tradition, which mostly develops in Philosophy of Education. I will illustrate some of the most prominent accounts in this tradition, and argue that the way they characterise indoctrination—as a degenerate form of proper education—is unsatisfying. Thus, after posing desiderata for what constitutes a more satisfying account of indoctrination—including developing what Shiffrin (2000) calls a motive-based characterisation of it—I will turn to the literature on manipulation, and particularly Yaffe's (2003) views.

In chapter 3, I will build upon Yaffe's conception of indoctrination. First, I will assume his view that indoctrination is a form of manipulation where an indoctrinator manipulates his subject's pattern of response to reasons so that she systematically chooses to do his bidding. Then, I will show that this sort of manipulation causes

the emergence of a defective pattern of reasoning in the subject, one that is self-preserving and self-reinforcing.

Further, I will show that the emergence of this pattern is not unique to indoctrination, and I will provide two examples: echo chambers and—what I call—sticky beliefs. From this comparison, I will infer that—from a viewpoint of doxastic cognition—indoctrination is not a kind of its own, but rather a manifestation of a more general epistemic phenomenon, namely the emergence of the defective pattern. However, I will argue, while in echo chambers and with sticky beliefs the pattern emerges spontaneously, in indoctrination its emergence is sought by the manipulator, because it suits him. Hence, I will conclude, indoctrination is best characterised as the exploitation of an epistemic-cum-cognitive vulnerability of the mind.

In chapter 4, I will explore a practical application of my epistemological account. That is, I will address the question of how to counteract indoctrination. I will argue that if we assume that indoctrination is bad education and we try to supplement it with proper education, our attempt at counteracting it backfires because it generates belief polarisation. Instead—under the assumption that indoctrination is the instilling of a defective pattern of reasoning—we should aim at bypassing the pattern by influencing the subject's feelings and making her aware of her epistemic practices.

In the concluding remarks I will highlight why I believe that having conducted a descriptive analysis of indoctrination will have led me to a more explanatory account than those in the Analytic tradition. Above all, I will argue that conceiving indoctrination as a form of manipulation, and not of bad education, solves various issues that made the accounts in the tradition unsatisfying.

Chapter 1

Proposing an approach

1.1 Why an epistemological account?

What is the point of developing an *epistemological* account of indoctrination? In a nutshell, the epistemological one is a unique perspective because it offers an insight into all instances of indoctrination at once, regardless of the context.

Indoctrination, as we commonly understand it, has various manifestations. Among the most famous ones, there are the authoritarian regimes of the 20th Century, such as the Third Reich and Mussolini's regime. There are what—in popular culture—we call cults and sects, such as the Peoples Temple and the Manson Family. There are some large, powerful churches, such as Jehovah's Witnesses and The Church of Jesus Christ of Latter-day Saints (Mormon Church). There are terrorist groups that groom potential members to recruit them, such as the various jihadist movements. And then there are organisations the sociological classification of which is trickier, but where we can confidently say indoctrination occurs, such as Scientology and big cats sanctuaries in America.¹

So, indoctrination is a pervasive phenomenon, one which permeates and shapes all sorts of environments—religious, political, private, online, etc.—and operates on different scales of magnitude—ranging from global organisations to close-knit communities. For this reason, instances of indoctrination differ from one another in

¹That is, the deeply interconnected society of big cat conservationists and collectors in America in their private zoos and sanctuaries, that has been depicted in 2020 true crime documentary *Tiger King*. This includes Doc Antle's Myrtle Beach Safari and what used to be Joe Exotic's The Garold Wayne Exotic Animal Memorial Park.

various respects, and it is difficult to explain why they all are, on some deeper level, instances of the same phenomenon. Moreover, because indoctrination manifests itself in such a variegated manner, it can be difficult to draw the line between cases that are indoctrination and those that are not. My claim is that with an epistemological account we can overcome these complications.

The first way in which instances of indoctrination differ from one another concerns the practices they involve. Indeed, indoctrination can be carried out in various ways. Depending on the context, different techniques are employed, different dynamics involved, as well as different customs. Even though this does not trump our intuition that each instance of indoctrination boils down to the same deeper phenomenon, we cannot but notice that instances of indoctrination *look* different across contexts.

Take, for example, two cases that could not be further apart: indoctrination in the Third Reich and indoctrination in Doc Antle's Myrtle Beach Safari. In these cases, different social structures are involved, different techniques employed, different ends pursued, etc. The leaders of the Third Reich used to employ propaganda, youth organisations, control of the media and state censorship to spread an ideology of patriotism and racism, as well as silence political dissent. Indoctrination was a tool for the party in power to obtain the support of an entire population, and it was carried out through ad hoc, widespread public institutions. Although the citizens of the Reich were the target of indoctrination, the average citizen would hardly have any contact with the people designing their indoctrination—e.g. the minister of propaganda, the party intelligentsia—and would more likely only have contact with those carrying out the plan—e.g. local functionaries.

By contrast, the Myrtle Beach Safari, a wildlife and safari park in Southern California, is an intimate, closed community, where the members—hired as voluntary staff—live under the direct influence of charismatic leader Doc Antle. According to testimonies recorded in Episode 2 of docu-series *Tiger King*, members of staff, mostly women, are selected for being young and inexperienced, introduced to the community, and then made to detach from their contacts outside the park.

Their identity is dissolved through name, dress and body modification and—under the influence of Antle—they develop a new work ethic which demands absolute dedication to the animals, as well as submission to the leader. This process of assimilation into the community is aimed at ensuring the successful running of the park and the gratification of Antle.

From a sociological point of view, although these cases share some similarities—such as the cult of ruthless leaders and the lifestyle-shaping quality of their doctrines—it is not clear what exactly makes them instances of the same phenomenon, i.e. indoctrination. An in-depth description of the mechanisms of propaganda etc. in the Third Reich detracts from an in-depth description of the practices of identity replacement etc. in the Myrtle Beach Safari. And even if the two descriptions were not competing, the connection between them appears feeble. However, assuming indoctrination has an epistemological component, the analysis of that component should be able to explain why both cases are instances of indoctrination: The practices may differ, but what goes on in the mind of the indoctrinated is the same.

It is not just the practices, however, that vary across contexts. It is also the harms that indoctrination produces, and the moral significance of those harms. We tend to think that indoctrination prepares the ground for abuse; that is, if carried out successfully, it makes the subject of indoctrination more vulnerable to abuse. And yet it is not obvious that indoctrination *necessarily* leads to abuse because it does not always lead to the same type of abuse.

Going back to the previous example, indoctrination in the Third Reich mostly raises questions that pertain politics. We morally object to that sort of indoctrination for reasons that have to do with human and civil rights, justice and democracy. We are concerned with the violation of the citizens' right to proper education, their right to public information, etc. We are concerned with the intuition that indoctrination undercuts the freedom of thought and effectively silences those who speak up. We are concerned with understanding how exactly indoctrination undermines justice and democracy, and how it reinforces the political power of those who already

have it. By contrast, when hearing testimony from the Myrtle Beach Safari, we are mostly concerned with the emotional, sexual and financial abuse the members are put through daily, the terrible living and working conditions they are convinced to settle for, and the trauma they experience when they leave the community.

As seen in these two cases, indoctrination produces a variety of harms, but none of these harms is produced necessarily. In general, we find indoctrination objectionable, and we understand it as morally wrong. But it is difficult to provide an explanation for its wrongness that covers all instances of indoctrination. If we simply ask why indoctrination is harmful or unjust, the answer will vary case by case. However—under the assumption that indoctrination has an epistemological component—if we ask about the epistemological harm of indoctrination, the findings can be applied to all instances of it.

The epistemological viewpoint has the further advantage of isolating the cases that are indoctrination from those that are not. This is especially important because there are plenty of ambiguous situations. For example, we might not want to say that members of incel communities online are indoctrinated in the same way sect members are, because these online communities lack a strict orthodoxy, an indoctrination programme, and a group of indoctrinators officially in charge of enforcing it. And yet, there are some striking similarities between the mindset of incels and sect members—such as the radicalisation into the values of their groups and the camaraderie between the members—which makes us wonder what exactly these groups have in common, if not indoctrination.

Similarly, we might not want to say that girls in a patriarchal society are literally indoctrinated into believing that they should be subordinated to their male counterparts. At least not indoctrinated in the same sense in which children are sometimes indoctrinated into sacred scriptures in a religious school. That is because girls—at least nowadays, in most liberal societies—tend to pick up these values passively, by being exposed to a value-laden language and imitating the behaviour of other women around them, while indoctrination into the sacred scriptures is typically carried out in the form of an education programme that is consciously taught

and learnt. And yet, the resemblance between these cases is uncanny: In both cases, the effect is the deep sinking in of particular notions and norms.

For a final example, some people become completely unrecognisable every time they reunite with their second family at their community centre or at the supporters club for their favourite football team. Indeed, the religious-looking aspect of association football culture is proverbial. That is because, in certain environments, the subculture develops in such a way that the members become obsessed with its symbols, vocabulary, and practices. Again, we should be hesitant to call this indoctrination, but it appears to have something in common with it.

These examples show that we are probably less competent than we think at spotting indoctrination when it happens. We are perfectly able to identify the paradigm cases of indoctrination: authoritarian regimes, apocalyptic sects, fundamentalist religious education, etc. But when we move away from these obvious cases and consider more ambiguous ones, we have a hard time classifying what is going on, and we may even mistake for indoctrination phenomena that only resemble it. Especially, it seems, we have difficulty separating indoctrination from other forms of human interaction which involve the transmission of information and values.

My claim is that if—as I assume—indoctrination has an epistemological component, and we identify it, then we should be able to tell instances of indoctrination apart from these other similar-looking phenomena, which presumably are not characterised by the same epistemological component.

To summarise, indoctrination takes various forms in terms of practices involved, it admits of various harms though none of which necessarily, and it shares a resemblance with other phenomena which we can easily mistake for indoctrination. So, to capture what indoctrination is about at its core, we need to encompass all the different manifestations of indoctrination. I claim that we can do that by focusing on the epistemological aspect of indoctrination. By developing an epistemological account, we gain a sufficiently deep insight into each instance of this multi-faceted phenomenon, while also demarcating what counts as an instance of indoctrination

and what does not.

1.2 A descriptive analysis

Which leads to a methodological question. When I advocate for developing an epistemological account of indoctrination, what sort of project do I mean exactly? Is it just about taking note of the common features shared by famous cases of indoctrination? That is not the idea.

The idea is to *discover* something about the cognitive mechanisms at the heart of indoctrination, i.e. what I previously called its epistemological component; the aim is to uncover and shed light on whatever is going on in the mind of indoctrinator and indoctrinated. To do that, the best way to conduct the inquiry is—I believe—to do a *descriptive* analysis of indoctrination, as opposed to a *conceptual* analysis, an *ameliorative* analysis, or a *genealogical* analysis.

According to Haslanger (2000), usually the philosophical analysis of a concept falls under one of these four categories: descriptive, conceptual, ameliorative and genealogical. Each category corresponds to a different methodology and a different understanding of what it means to analyse a concept. Correspondingly, there are four ways to answer the question: Which philosophical projects count as analysing a concept? Here are in brief the four possible answers conceived by Haslanger.

A descriptive analysis of concept *x* consists in asking what *kind*, if any, our epistemic vocabulary—in this case, the term ‘*x*’—tracks, and developing a more accurate concept of it. To conduct a descriptive analysis of *x*, we first identify the paradigm cases of *x*. We then observe these cases closely—usually by relying on empirical or quasi-empirical methods²—to explicate the kind that the paradigm cases belong to. Finally we update our concept of *x*. The final product of a descriptive analysis is thus increased empirical knowledge of the kind which our term ‘*x*’ tracks—which is also to say a more refined concept of *x*.

A conceptual analysis of concept *x* consists in asking what is *our* concept of *x*, i.e. what we have in mind when we speak of *x*. Since what we have in mind might

²For example, the descriptive analysis of *epistemic* concepts can rely on cognitive science; the descriptive analysis of our concepts of ‘water’ and ‘rainbow’ can rely on Physics and Chemistry.

be blurry, conceptual analysis is meant to make it clearer. When conducting such an analysis, we refine the concept of *x* until we identify the conditions under which a competent speaker would say that something is *x*. Conceptual analysis is conducted through a-priori methods. Because the object of our analysis is a mental content, we access it through introspection. By registering our intuitions about *x*, and weighing them against each other, we succeed in the conceptual analysis of *x* when we reach a reflective equilibrium.

An ameliorative analysis of *x* is conducted by asking what the *point* of having concept *x* is, and how we can improve our concept so that it serves its purpose better. So, to ameliorate a concept is to assign it a new meaning, which reflects the role we want the concept to play within a theory. In turn, the theory should be aimed at driving concrete progress and improving society.

Finally, a genealogical analysis of concept *x* consists in exploring the *origins and history* of *x* to understand how it is embedded in our evolving social practices. To conduct such an analysis we study which behaviours, rules, institutions, etc. are described by *x*, but also are structured by *x*. Crucially, all these practices surrounding concept *x* are, on the genealogical approach, what *x really* is.

When it comes to the concept of indoctrination—and keeping in mind that the aim is to develop a distinctly epistemological account—I believe a *descriptive* analysis is the most interesting one to conduct, because it has the greatest potential to improve our understanding of indoctrination. Indeed, given our limited knowledge on the topic, it is too early to conduct any other kind of analysis.

Our intuitions as to what happens psychologically when indoctrination occurs are too feeble to motivate a *conceptual* analysis. Sadly, we often characterise the cognitive mechanisms of indoctrination as brainwashing and mind control, and only have insight into some of the techniques employed by indoctrinators. Conceptual analysis is interesting, when *our* concepts are vivid enough, which is not the case here.

Our vague understanding of indoctrination may seem a good reason to embark on an *ameliorative* analysis, especially given how unhelpful the myth of brainwash-

ing is, as there is a tendency to use this notion to indiscriminately insult education and communication practices we disapprove of. An ameliorative analysis of indoctrination from an epistemological point of view would serve the purpose of redefining our understanding of the psychology of indoctrination, to then locate it into a wider frame, and compose a picture of psychological mechanisms that can be used to improve society.

But while amelioration has inspiring goals, our current understanding of the psychology of indoctrination is again too poor for us to substantially re-define it and have it serve our practical purposes. We are still at the beginning of understanding how indoctrination works, so the most pressing task for us now is to discover more about it, in the same way in which a scientist would go about and uncover the mechanisms of nature.

A *genealogical* analysis would seem to defeat this purpose. Arguably, indoctrination is a peculiar notion, in that we know more about the social practices surrounding it than how it works. Indeed, the history of the concept is well known, as well as its current characterisation as brainwashing and usage in dystopian literature and political rhetoric, but we do not yet understand the cognitive mechanisms that characterise it. A genealogical analysis of indoctrination is of course important, but we also need an explanation of these mechanisms, as part of a more organic understanding.

All in all, a *descriptive* analysis is the most appropriate and the most needed now to increase our knowledge of the epistemology of indoctrination. We should identify and closely observe the paradigm cases of indoctrination, explore which cognitive mechanisms are at work in those cases, and re-define our concept of indoctrination by drawing on our discoveries. If we are successful, we will be able to make a claim such as the following: Of all features indoctrination, *this* particular mechanism is the most striking, relevant one; *this* particular mechanism is the closest phenomenon to what we take ourselves to be referring to when we talk about indoctrination. *This* is what determines the membership of the paradigm cases to the same kind. Therefore, we should re-conceptualise indoctrination as being es-

entially characterised by *it*.

Having stated my suggestion as to what constitutes a promising approach to understanding indoctrination, in the next chapters I will develop an account of indoctrination following my own advice, to show that what I propose is indeed a fruitful strategy.

To conduct my descriptive analysis, I will first dig in the philosophical literature and the paradigm cases for clues as to what cognitive mechanisms characterise indoctrination. Theoretically, such an analysis could lead to surprising findings. This is fine: We are not conducting a conceptual analysis, so we can allow for the findings to differ from our intuitions and potentially overthrow them.

Chapter 2

Drawing from the literature

2.1 Indoctrination in the literature

Focusing on the literature on indoctrination in the Analytic tradition, we find that the discussion around indoctrination predominantly—almost exclusively—develops in Philosophy of Education and Legal Philosophy, and even in the latter case, it is strongly interconnected with the topic of education. Throughout the literature, indoctrination is conceived of as intrinsically connected to the concept of education, and it is assumed to be a bad kind of education, i.e. proper education somehow falling short. Unified by this common thread, the literature can be divided into two main approaches: descriptive and normative.

The normative part is borrowed from a narrower debate on the rights of parents and children in child-rearing, especially in a religious context. What interests us in this debate is the question about the legitimacy of indoctrinating children, and especially one's own. Here, authors essentially weigh the rights of the children against those of the parents, to determine if and when religious indoctrination is justified.

For instance, Feinberg (2007) famously proposes the principle of the child's right to an open future, according to which parents should not make decisions that undermine the future autonomy and capacity of self-determination of their children. Among the harms that parents should avoid inflicting, bad education and a lack thereof rank the highest, and the influence of religion on education poses a particular

threat. In light of this principle, religious indoctrination is illegitimate insofar as it violates the child's right to an open future.

Thinking along similar lines, legal philosopher Mahoney (2014) addresses the question of how to deal with religious objections to education policies and argues that we should accommodate them only if we can uphold adequate education standards. In direct response and as an objection to Feinberg, Mills (2003) argues that an open future is impossible and undesirable, and claims that it is unreasonable to expect parents to encourage their children to pursue lifestyles and choices that violate the parents' moral principles. Reflecting on her own experience, Overall (2010) maintains that "parents are not morally culpable for indoctrinating their children if they lack the background, education, and cultural context that would enable them to evaluate the justifications for their actions" (p. 11). Finally, Montague (2000) argues against the existence of parental rights, including the rights of parents to decide how their children should be educated and which religion they should be encouraged to accept.

The normative side of the literature is not very helpful for developing an epistemological account of indoctrination, as it consists in an insular debate on a specific application of our thoughts on indoctrination. Questions about child-rearing are but a small subset of the philosophical concerns which indoctrination raises, so we should not treat them as the central point of reference in this project. By contrast, the descriptive side of the literature calls for closer inspection and careful consideration, because it addresses a crucial question: It seeks to understand what indoctrination is, and in some cases how it works. Although I will later argue that my epistemological account is a departure from this tradition, this is the literature my project is in dialogue with.

Condensed in a collection of essays edited by Snook (2010a), this part of the literature consists in an ongoing debate between philosophers of education who "are concerned primarily with the meaning of the term 'indoctrination'" (Snook, 2010a, p. 2). In the light of Haslanger's taxonomy, we can consider this endeavour as a collective effort at conceptual analysis.

While the strategies employed by the various authors are diverse, the common element between all accounts is an underlying assumption about the nature of indoctrination, i.e. the idea that indoctrination is a degenerate form of teaching, differing from proper education in one or more of three elements: method, intention and content. Method consists in the techniques employed for teaching, intention is the outcome desired by the teacher, and content is the subject matter being taught. I call this the tripartite model.

Further, under this model, there is consensus that “indoctrination must involve the handing on of beliefs as distinct from skills, attitudes or ways of behaving” (Snook, 2010a, p. 2), and Snook calls this a first restriction on content. That is, all authors agree that indoctrination has to do with the transmission of beliefs, reasons, ideas, etc. as opposed to know-how, habits, and emotional response to stimuli. As Wilson (2010) explains: “Roughly, if I illegitimately (whatever this may mean) persuade a child to think that God will punish him for masturbating, this is indoctrination: if I simply give him a feeling of fear and repulsion about it, this is conditioning” (p. 14).

What is astonishing is the variety of conflicting views under the tripartite model. The debate is characterised by disagreement on various points, most notably which *criteria*—among method, content and intention—we should use to define indoctrination, and which particular methods, contents and intentions are the *relevant* ones. There is also disagreement on which instances of teaching count as indoctrination, and whether indoctrination is necessarily wrong.

Crucially, it appears that under this model it is difficult to formulate an account of indoctrination that establishes necessary and sufficient conditions for the occurrence of indoctrination. For any specification of relevant method, content, or intention, it seems possible to cast doubt on that formula by providing a sufficiently convincing counterexample, or an equally convincing alternative account.

Among the writers who think that indoctrination consists in the use of certain *methods* of belief-transmission, it is a popular view that indoctrination necessarily involves teaching by *non-rational* methods. A defender of this view, Wilson (2010)

argues that indoctrination is the implantation of beliefs that are causally motivated, for instance motivated by the desire to obey an authority figure, as opposed to rationally motivated, which is typical of proper education. Similarly, Atkinson (2010) argues that what indoctrination lacks, but proper education provides, is “adequate support, by way of proofs, reasons, evidence” (p. 44). He also compares education to training and indoctrination to drilling, stating that the trained man knows not only how to proceed, but also why he should proceed that way, while someone who is drilled can only mechanically apply rules of thumb.

A different way to defend the method criterion is to characterise indoctrination as *authoritarian* teaching, as opposed to democratic teaching, a title reserved for proper education. This restriction of method reflects the history of the word ‘indoctrination’. As Gatchel (2010) recounts, until the 17th-century ‘indoctrination’ and ‘education’ referred to the same practice, but with the rise of some concepts of democracy in the United States, the democratic discourse appropriated itself of the concept of education, attributing to it a new, liberal and progressive connotation (especially referring to individual growth), as opposed to the old, authoritarian and reactionary practice of indoctrination.

What qualifies as authoritarian teaching is of course debated. Moore (2010) suggests that authoritarian education—i.e. indoctrination—is one which stresses data and facts while neglecting evidence or justification for these, so that the pupil—not being used to it—will not “some day ask for reasons where it is unwise or impossible to give them to him” (p. 76). Interpreting ‘authoritarian teaching’ differently, Kilpatrick (2010) argues that indoctrination is characterised by the anti-democratic technique of imbuing children with one’s personality and views while they are still docile, ignorant and have little capacity for critical resistance.

While a constraint on method may be appropriate, there is a problem in considering it the *sole* criterion of indoctrination: Whatever method we may pick, it can be argued that it is not exclusive of indoctrination. For a classic example, the method of drilling is typically used to teach the multiplication tables to young children, and yet it is arguable that that is not a case of indoctrination. For instance, a

defendant of the content criterion could argue that you cannot indoctrinate into the multiplication tables because they are self-evident truths.

For a more complex example, consider a school where History of Philosophy is taught by imparting inert ideas,¹ that is by reducing “both method and subject-matter to the status of an object of passive acquisition” (Garrison, 1986, p. 267). In this school, inert ideas are poured in the students’ minds just for the sake of it. This is anything but stimulation of the students’ rationality and intellect, but it is not a case of indoctrination either. One way to argue this is to say that you can only indoctrinate into a specific philosophical doctrine and not into the History of philosophical doctrines.

As for authoritarian methods of teaching—a suspicious category to begin with, for ‘authoritarian’ is often more of an evaluative judgement than a description—they also seem to be employed in practices other than indoctrination. Typically, we instil children with notions of hygiene and manners, we teach them their mother language, and we cannot avoid transmitting them at least some of our values. We do not normally give them reasons or justifications for these teachings, nor do the children have much capacity for critical resistance, and yet it can be argued that these practices should not count as indoctrination. For instance, a defendant of the intention criterion could say that these cases are not indoctrination because they are inevitable and motivated by good intentions.

If method alone is not very promising as a defining criterion of indoctrination, content-based and intention-based ways of defining indoctrination are just as unconvincing. When it comes to *content*, various constraints have been proposed. It is frequently suggested that for indoctrination to occur, the subject matter of the teaching should be some kind of *ideology, doctrine, or political agenda*. Flew (2010a,b) is a proponent of this view. Though not a defender of content as the sole criterion, he argues that indoctrination is the attempt to implant doctrines that are either false or not known to be true. He argues that religious teaching must consist in indoctrination because religion does not constitute knowledge, and says that the teaching of

¹A coinage of Alfred North Whitehead.

morality is indoctrination insofar as *oughts* are presented as facts. Additionally, he concedes that one can be indoctrinated into the truth, so long as the subject matter is intertwined with a doctrine. For instance, the teaching of French could constitute indoctrination only if “tied up with something wider and more ideological” (Flew, 2010a, p. 55) such as a political-racial agenda.

If Flew’s explanation of what constitutes a doctrine is vague, Gregory and Woods (2010) provide a more thorough account. They claim that the content of indoctrination must be a doctrine, and that doctrinal beliefs are different from ordinary beliefs in two regards. First, they are intimately tied up with actions: Accepting them entails “commitment to act in particular ways, to profess and act out a particular value and way of life” (Gregory and Woods, 2010, p. 129). Second, there is something about doctrines that causes each doctrine to be “seized upon and elevated to the status of a universal truth . . . against which nothing can count” (Gregory and Woods, 2010, p. 130).

Contrary to Flew, Gregory and Woods maintain that it is not possible to indoctrinate into truths. They write: “It is virtually impossible to make sense of someone who says that he intends to get a boy to believe unshakeably that big ben is in London” (Gregory and Woods, 2010, p. 142). So, though Flew disagrees with Gregory and Woods on whether truths—provided they are tied up with an ideology—can be taught through indoctrination, they all otherwise agree that indoctrination consists in the inculcation of doctrines.

By contrast, White (2010), maintains that it is wrong to restrict the content of indoctrination to ideologies because, when we worry about indoctrination, we do not have just ideology in mind. He says: “What of those schools where in a hundred and one different ways some teachers try to get their pupils to see themselves as future hewers of wood and drawers of water?” and “what, too, of teachers who try to fix in some of their pupils’ minds the ineradicable belief that they are of limited intelligence?” (White, 2010, p. 96).

He further asks us to imagine a teacher trying to indoctrinate a pupil into the belief that Melbourne is the capital of Australia, i.e. a false belief the content of

which is not ideological (religious, political, etc.). This, he thinks, is a logical possibility; but he recognises that the defenders of ideology as content restriction may object that such indoctrination is impossible unless the teacher backs up their claims by inculcating further notions into the pupil, e.g. that talking to Australians is wrong. But this, says White, far from being a restriction on content, is a restriction on method, i.e. the method of backing-up beliefs with other beliefs.

So, like the method criterion, the content criterion is controversial and the object of radical disagreement. As Snook (2010a) summarises, “some [authors] hold that any beliefs can be indoctrinated; others argue that ‘indoctrination’ can be used only of doubtful or false beliefs; a third group argue that only doctrinal beliefs are subject to indoctrination” (p. 2).

The *intention* criterion does not fare any better, as authors are far from reaching consensus about which intentions of the teacher, if any, are relevant for indoctrination. Among the defenders of the intention criterion, there is the editor of the collection of essays, Snook (2010b). To identify the relevant criterion, Snook implements an original strategy. He compiles three lists of examples: “cases which are clearly indoctrination”, “cases which may seem like indoctrination, but which are not since they are unavoidable”, and “problematic cases” (Snook, 2010b, p. 118). He then considers, in turn, the three criteria—intention, method and content—and checks which of them can explain why each case belongs to its list. He also repeats the procedure with all combinations of the criteria, including the conjunction of all three criteria at the same time.

His conclusion is that “only intention can serve as an adequate criterion for distinguishing indoctrination from education, and attempts to link intention conceptually with another factor such as content or method will destroy the delicate balance” (Snook, 2010b, p. 119). By intention, he means “what is desired” and “what is foreseen as likely or inevitable”, but not “what is foreseeable”.²

As for the relevant intention, Snook says that “a person indoctrinates P (a

²The idea is to rule out cases in which the teacher is so naïve that he thinks he is teaching his pupils properly, but in fact he is accidentally indoctrinating them. This, Snook believes, should not count as indoctrination.

proposition or set of propositions) if he teaches with the intention that the pupil or pupils believe P regardless of the evidence” (Snook, 2010b, pp. 119–120).

This is similar to what most defenders of the intention criterion say. Indeed, most agree that the relevant intention has to do with *fixing beliefs beyond recall* in the pupils’ minds. For instance, White (2010)—whose argument against the content restraint I cited earlier—agrees with Snook that indoctrination is definable only in terms of intention, the relevant one being to make the pupil “believe that a proposition ‘p’ is true, in such a way that nothing will shake that belief” (White, 2010, p. 93). The same is maintained by Wilson (2010)—who we already saw defending a constraint on method.

Despite its popularity, this idea is the target of an objection that is hard to escape. According to Crittenden (2010), who favours a method criterion:

It makes no difference for the concept of indoctrination whether the teacher is in good or bad faith. A teacher who is not personally convinced by valid arguments and yet presents them as though he were is not indoctrinating, while a teacher who thinks he has adequate evidence for certain beliefs and proposes it as such is indoctrinating, if in fact he is mistaken. (p. 113)

More generally, it is safe to say that whatever intention we take to be the relevant one, there are cases in which we intuitively think indoctrination occurs and the relevant intention is missing. This fact undermines the intention criterion at its root.

This objection seems to be countered by Wilson’s (2010) argument that “indoctrination is not wholly to be defined by the *conscious* aims of the indoctrinator” (p. 15). He says:

To be an indoctrinator, a person must certainly intend his pupil to arrive at a certain belief, but he need not specifically intend that the pupil should always maintain the belief in the face of reason, or that he should reach it as a result of bowing down to the indoctrinator’s authority, or anything else of that kind. (Wilson, 2010, p. 15)

Wilson's point is that what people say to themselves that they are doing is not the same as what they are in fact doing, and so it is possible to indoctrinate without knowing that you are.

But if this should count as a defence, it is weak, for if intention is the criterion of indoctrination, how is it possible for someone to indoctrinate without knowing that they are? To grant this possibility, we would have to concede that the relevant intentions for indoctrination can be unconscious. For example, even if a teacher believes that he intends to cultivate the pupils' critical thinking and takes all the measures to succeed, we should consider that indoctrination insofar as the relevant intention lives deep in his heart, concealed even to him, and is in fact the force that moves him and informs his action.

The problem with this defence is not only that it commits us to the existence of these unconscious intentions and a precise way in which they would work. The real problem is that if we take unconscious relevant intentions to be the trademark of indoctrination, then it becomes impossible for anyone to determine whether indoctrination is ever going on. Crucially, indoctrination would become *de facto* indistinguishable from proper education. Which in turn would make it mysterious why we should prefer proper education over indoctrination, thus contradicting the assumption that indoctrination is the degenerated, undesirable version of proper education. So, defending the intention criterion by appealing to unconscious intentions is at best useless, and at worse inconsistent.

Snook's argument in favour of the intention criterion is also unconvincing, because it is self-referential. The circularity stands out when we conceptually divide his strategy into two phases that feed into each other. First, we see him rank and classify cases of suspected indoctrination based on how sure he is that they are indoctrination. And then, we see him proceed to uncover what his personal criterion was for ranking and classifying them that way in the first place. In doing so, Snook conducts a conceptual analysis of his personal concept of indoctrination. However, not everyone conceives indoctrination this way, as we have seen, and another writer would have compiled the lists of cases differently, or listed a different set of cases,

and drawn a different conclusion. So, while Snook's intention-based definition of indoctrination is coherent and *prima facie* attractive, it does not establish that this is *the* concept of indoctrination that is, in general, currency in the discussion.

Having exposed the weakness of the intention criterion, as well as those of method and content, it seems reasonable to question the utility of the tripartite model. The idea that a combination of these criteria could constitute necessary and sufficient conditions for the concept of indoctrination is initially appealing but does not look feasible. Whatever criteria we assume, the remaining criteria can be employed to construe a plausible counter-example: For instance, where someone suggests the intention criterion, someone else can respond with an example of something that intuitively seems like a case of indoctrination, but where method and content, rather than intention, seem to be the features that make it a case of indoctrination. And even where a criterion is agreed upon, there is disagreement on the particular constraints that should be posed.

A further motivation for questioning the tripartite model is the suspicion that, under this model, we are only apparently discussing three substantially different criteria. The tripartite model gives us the illusion that we are capturing the essence of indoctrination when picking the correct combination of criteria. But intention, method and content could otherwise be seen as different, interchangeable lenses through which we look at the same phenomenon. For example, imagine that we have the strong intuition that indoctrination is intimately tied with the notion of unshakeable beliefs. Under the tripartite model, we ought to choose whether to transform this intuition in a method constraint, an intention constraint or a content one. But whatever we choose, does it make a significant difference? Probably not, given how sometimes authors in the literature disagree on which label—'content', 'method' or 'intention'—to give to a particular constraint.

I suggest that instead of subsuming our intuitions under fuzzy labels, it would be more productive to discuss our intuitions directly. For instance, if we think that indoctrination has to do with unshakeable beliefs, then we should analyse what

these beliefs are, how they work, what they do, etc.³ In doing so, we may find out that method, content and intention play indeed some role in indoctrination, but this role is more marginal than the tradition makes it.

All in all, it appears that the assumptions of the tripartite model interfere with the direct analysis of the essential features of indoctrination; in particular, two assumptions are questionable. Namely, that indoctrination is a form of teaching, i.e. the degenerate version of proper education. And that indoctrination must be defined in terms of content, method or intention. By dropping these two assumptions, we depart significantly from the philosophical current that characterises the contemporary debate over the definition of this concept.

2.2 Beyond the tripartite model

Going beyond the tripartite model, we ought to reconsider what a satisfying account of indoctrination looks like. Up to this point, I have committed to developing a descriptive analysis of indoctrination, with a focus on its epistemological aspect. But where to start doing that, what are the basic rules, the minimum desiderata? I pose two.

The first minimum desideratum is that an account of indoctrination should be developed thoroughly enough to shed light on some important mechanisms of indoctrination, such that indoctrination occurs if and only if these mechanisms are in place. In other words, the account should be a tool sharp enough to draw a neat line between what is a case of indoctrination and what is not, by providing a unique criterion, which is satisfied when a precise cognitive mechanism is at work in a given situation.

There may be more than one way to satisfy this desideratum. There may be more than one cognitive mechanism—or a set of them—that could be considered a good criterion of indoctrination, and correspondingly more than one account that

³The only author proposing a view of this sort is Green (2010), who argues that our belief systems are anything but logical, because beliefs can be held without evidence and even against evidence, and indoctrination has to do with how beliefs are held, i.e. non-evidentially. This is an odd consideration to make under the tripartite model, as it has to do more with the *style* with which beliefs are held, than any of the three criteria.

captures it. This is especially true if we slightly shift our viewpoint, for instance from pure Epistemology to Behavioural Psychology. I do not mean to develop the only possible correct account of indoctrination, only a convincing one. That my account should identify and mark out the most significant features of indoctrination is commanded by the principles of descriptive analysis; the additional desideratum—that the account should be strong enough to decide whether any situation is a case of indoctrination—is perhaps ultimately a way of doubling down on the commitment to descriptive analysis.

Any account under the tripartite model is far from satisfying the first desideratum, as it is insufficient to distinguish cases that are indoctrination from the ones that are not, and insufficient to provide a general explanation as to why something counts as indoctrination at all—including explaining the difference between indoctrination and proper education. Posing the first desideratum is a resolution to solve these issues.

The second minimum desideratum is to develop a type of account that Shiffrin (2000) calls *motive-based characterisation*. According to Shiffrin, we can account for certain phenomena by aiming at explaining their normative significance to us. She uses this strategy to account for paternalism:

One plausible approach, I would suggest, involves not merely testing formulas against intuitions, but also testing formulas with an eye to arriving at a conception of paternalism that fits and makes sense of our conviction that paternalism matters. That is, it seems worthwhile to assess what is central in our normative reactions to paternalism and to employ a conception of paternalism that complements and makes intelligible our sense of paternalism's normative significance. (Shiffrin, 2000, p. 212)

Like paternalism, indoctrination lends itself well to this sort of account, because 'indoctrination' is a value-laden term, which not only describes a phenomenon, but also registers a negative evaluation of it. We find indoctrination morally objectionable, and aiming at explaining why seems like a promising way to

account for it.

Shiffrin's approach may seem incompatible with the plan of conducting a descriptive analysis—which is what I am doing. A descriptive analysis is meant to be a quasi-empirical inquiry. It is meant to look into the phenomena as they unfold in the world, not register our concepts, intuitions and feelings about them. That would be the objective of a conceptual analysis. So, one might ask whether I can legitimately adopt Shiffrin's approach. Ahlstrom (2008) shows why it is not only legitimate, but also necessary.

In his paper *Epistemology and Empirical Investigation*, Ahlstrom takes a stance on what epistemological investigation looks like—a stance similar to Haslanger's on what counts as a philosophical analysis for a concept. He claims that epistemological investigation is substantially empirical, and that is because—even if we can make successful reference to a kind, and even if the kind is artifactual—it does not mean that we have insight into all its properties, which can only be uncovered via empirical investigation.

So far so good. Ahlstrom's argument is relevant to our case, and he seems to be advocating for the importance of descriptive analysis, i.e. studying the actual phenomenon of indoctrination as opposed to our concept of it. He can be taken to say that a conceptual analysis of indoctrination can only reveal so much because there are properties of indoctrination which we can know only empirically—even if we are perfectly familiar with the concept and its application conditions.

However, Ahlstrom argues further, paying attention to our concepts is also important, because it aids our epistemological investigation. More precisely, we should always be aware of the *purpose* of our concepts, because that allows us to know which is the correct referent of a concept. For this reason, when we have a concept for a certain purpose and its referent does not fit the bill, we can either alter the concept or find an alternative referent.

If we follow Ahlstrom's line of reasoning, we can conclude that a motive-based characterisation fits well in a descriptive analysis of indoctrination. Paying attention to our notion of indoctrination and reflecting on what we find objectionable about

it serves the purpose of identifying and marking out the phenomenon in the world that makes the most striking referent for our term ‘indoctrination’. Ultimately, in keeping with the principle of descriptive analysis, this should allow us to *refine*—as opposed to clarify—our concept of indoctrination.

As we take on board Shiffrin’s methodology, we can predict that it will deliver where the tripartite model fails. Under the tripartite model, attempts at capturing the essence of indoctrination are carried out by registering inessential aspects of it: method, content and intention. Focusing on these aspects allows us to get a glimpse into what we find objectionable in indoctrination, but it does not leave us the room to analyse it in depth. By contrast, a motive-based characterisation revolves strictly around the normative significance of indoctrination. In depicting indoctrination this way, we aim directly to register why it matters to us, which is a more essential aspect than method, content and intention. Ultimately, both the tripartite model and Shiffrin’s strategy are successful if they can explain why indoctrination matters, but the former approach is dispersive, and the latter straight to the point.

Having posed two minimum desiderata, the next step in developing an account that satisfies them is to ask ourselves: What is it exactly that we object to in indoctrination? Why does it matter to us?

As argued in the introduction, the harms of indoctrination are many, and they vary along with the context; they range from undermining democracy to damaging the emotional integrity of the subject, and so on. Because of this, multiple accounts of indoctrination are possible. But if the goal is to develop an epistemological account of indoctrination, then it is worth focusing on the epistemic harms of indoctrination. In turn, it may be possible to identify a whole range of epistemic harms that indoctrination produces. Once again, it is a matter of marking out a precise sub-set of that range, a precise epistemic harm, that we take ourselves to be objecting to when we feel that indoctrination is epistemologically harmful.

The direction I shall take is to trust the intuition that indoctrination—from a psychological point of view—is best described as a form of manipulation. Indeed, when we imagine what goes on in the heads of the parties involved, we see an in-

doctrinator moulding, shaping and by various means influencing the behaviour and perception of his subject, in a way that leads the subject to advance the manipulator's interest as if they were her own.

For instance, when we think of the deeds of a successful cult leader, we recognise that he is surrounded by a group of people who put blind trust in him, follow him, worship what he says and are willing to do whatever he asks. Intuitively, the cult leader caused them to trust him blindly, subjected their will to his, and subtly conducted them to obedience, though without the use of coercion. If this characterisation of indoctrination is correct, it is easy to see why we can consider it a form of manipulation, and why we would morally object to it as such.

Following this intuition, we can look for discussion of the epistemic harm of indoctrination in the literature on manipulation—thus moving away from Philosophy of Education. That is, if we conceive indoctrination as a form of manipulation—which writers in this area of Philosophy also do (Garnett, 2015; Sripada, 2012; Yaffe, 2003)—it is possible to take on board some things that have been said of manipulation and apply them to indoctrination when appropriate.

First, we can look at the debate on how to explain the Manipulation Intuition. This is the strong intuition—on which there is consensus—that a person subjected to manipulation is not free, or that at the very least their freedom is significantly diminished when they are manipulated. There are two schools of thought debating why that is, and the disagreement originates in their different conceptions of freedom. Compatibilists say that manipulation undermines freedom because it inhibits the cognitive capacities of the subject; incompatibilists say it does because manipulated agents have no ultimate control over their actions. At a first glance, we could hope for this debate to decide for us what the specific harm of indoctrination is.

The ground on which compatibilists and incompatibilists mostly fight is the Manipulation Argument. The argument is put forward by the incompatibilists and is meant to show that compatibilists, by conceiving freedom as they do, are committed to a view that contradicts the Manipulation Intuition. As the argument goes, there is a contradiction between two compatibilist intuitions. First, that a manipulated

agent has the cognitive capacities which compatibilists usually consider relevant for being free. Second, that a manipulated agent is not free (Manipulation Intuition). So, either the compatibilist drops the Manipulation Intuition, or they drop their conception of freedom.

To this challenge, various compatibilist defences have been put forward. Sri-pada (2012), for example, claims that manipulation cases, far from posing a problem for compatibilism, justify it, because they provide evidence for compatibilist principles: They show that the manipulated agent does not have the relevant cognitive capacities for freedom. To prove his point, Sripada shows that in a series of experiments that he conducted on a population the belief that an agent is not free varies along with (i.e. is fully explained by) the belief that they suffer from corrupted information and deep self-discordance. This means, he thinks, that people generally have the intuition that an agent who has been manipulated has significantly damaged cognitive capacities, and so the first premise of the Manipulation Argument is false.

For another example, Garnett (2015) employs a different strategy to defend compatibilism from the Manipulation Argument. He proposes a new “form of compatibilist freedom, social autonomy” (Garnett, 2015, p. 93), which is defined as freedom from foreign wills. Thanks to this new notion, he is allowed to maintain at the same time that the cognitive capacities of a manipulated person are not significantly damaged (only interfered with), and yet the person is not free—because they are not free from a foreign will.

Does this debate help us in our quest to identify the most significant epistemic harm of indoctrination? Realistically, there will be no decisive victory in the debate anytime soon. If anything, because it is not clear what could settle the question. After all, the classical problem of compatibilism—of which this is a variation—is a notoriously difficult one, so much that Kant included it among the antinomies of reason, i.e. those pairings of thesis and antithesis the conflict between which empirical reason cannot settle.

Nevertheless, there is something to learn and borrow from the debate over the

Manipulation Argument. First, if the Manipulation Intuition is true, and we conceive indoctrination as a form of manipulation, then it is plausible that the Indoctrination Intuition is also true: Indoctrination undermines the freedom of the subject, or at least it limits it significantly. Second, while there may be no agreement as to why indoctrination is freedom-undermining, the question arises: Is it because being indoctrinated makes you less autonomous? Does the subject of indoctrination not own their actions and decisions? Third—though they disagree on the significance of it—compatibilists and incompatibilists alike agree that the cognitive capacities of the subject of manipulation are interfered with. We can thus assume this of indoctrination too.

So, within the context of the literature on manipulation, we get a picture of indoctrination that suggests at various points why indoctrination matters to us morally: It is a freedom-undermining practice, that at least threatens the autonomy of the subject, and surely somehow meddles in their cognitive capacities.

2.3 Yaffe's account

The second and most substantial takeaway from the literature on manipulation is what we learn from Yaffe's (2003) account. In his essay *Indoctrination, Coercion and Freedom of Will*, Yaffe identifies indoctrination as a form of manipulation, where successful manipulation occurs when the subject systematically chooses to do the manipulator's bidding.

To explicate how indoctrination works, Yaffe presents a fictional example involving David Koresh. A subject has been singled out by the minions of Koresh, has been lured to his compound, and at the end of a weekend "seminar" she believes him to be the messiah. Later on, when Koresh asks her for a donation, after careful reflection she accepts, because she believes that Koresh is the messiah.

In the example, Yaffe argues, the subject has *chosen* to make a donation, but she is also subjected to a force that *makes* her believe she should; that is, she has been manipulated. She sees herself as having reasons to donate. However, her compliance has been obtained by suspicious means.

According to Yaffe, we can obtain the compliance of a subject without force by causing her to respond to reasons in a way that is advantageous to us. In other words, we can shape her *pattern of response to reasons*. This can be described as:

In part a pattern of belief about which features of oneself and one's environment really are reasons for acting in particular ways, and in part a pattern of desiderative attitudes, or affective states, that move one to act in ways consistent with one's beliefs about which facts are reason-giving. (Yaffe, 2003, p. 342)

Under this view, indoctrination is a manipulation of the subject's reasoning. More precisely, it is the production of a pattern of response to reasons that is convenient for the manipulator.⁴ The successful manipulator determines which facts the subject recognises as reasons, and what she takes them to be reasons for. As a result, the subject chooses to do the manipulator's bidding, by carrying out his will as if it were her own.

According to Yaffe, indoctrination is not to be confused with another form of manipulation: coercion. Like indoctrination, coercion tracks the subject's compliance. Here, too, the subject sees herself as having reasons to act the way the manipulator wants. But unlike indoctrination, coercion does not alter the subject's pattern of response to reasons. Instead, it relies on the subject having an ordinary pattern, and gives her reasons to comply. For example, an armed robber can coerce the bank employee to open the safe by threatening her. In doing so, he is not meddling with her reasoning; he is giving her a reason to do as he says, and predicting that she will. In a nutshell, "sometimes we manipulate the way an agent responds to reasons, sometimes we manipulate what reason she has" (Yaffe, 2003, p. 340).

Patterns of response to reasons do not only change as a result of indoctrination. Neutral causal forces can alter them as well. In general, our experiences alter the dispositions and abilities through which we recognise and respond to reasons. For

⁴As I will argue at the end of section 3.1, the pattern of response to reasons produced by the manipulator does not always benefit him, but it is *conducive* to benefits for him. That is, the defective pattern may backfire and damage the indoctrinator, but if it performs well, that is to the advantage of the manipulator.

this reason, neutral causal forces can cause a person to exhibit the same pattern of response to reasons as a subject of manipulation. Consider Yaffe's example. There are two people, Manipulated and Unlucky. Manipulated has been indoctrinated by Koresh, Unlucky relies on tasseography in her decision making. At one point, Koresh asks both to make a donation to him. Manipulated decides to accept because she believes that he is the Messiah; Unlucky accepts too because her tea leaves say that Koresh is trustworthy.

The difference between indoctrination and neutral causal forces is that indoctrination not only shapes our pattern of response to reasons, it also *crystallises* is. Once the manipulator has produced a pattern that suits him, he takes steps to make sure that it stays that way. Any time the subject seems to derail from the desired course, the manipulator intervenes to place her back. This explains why Unlucky may decide to donate to Koresh today, but not the next time he asks: She may lose faith in tasseography in the meantime, and acquire a new pattern of response to reasons that tells her to distrust Koresh. By contrast, Manipulated faces a significant obstacle in losing her faith in Koresh due to his lingering influence. Her pattern of response to reasons, which serves his ends, is constantly monitored and corrected by him. So, it is more unlikely to change.

To summarize, according to Yaffe indoctrination is a form of manipulation where a manipulator shapes how his subject responds to reasons, so that she will almost certainly choose to do his bidding. Additionally, the manipulator tracks the subject's compliance: Every time she is on the verge of changing the way she responds to reasons, he intervenes to prevent it.

Yaffe's account of indoctrination is so striking, that I will take it in full as the foundation for my epistemological account. By accepting what Yaffe posits, I find myself in the favourable position to say that my account already satisfies both the minimum desiderata, because that is what Yaffe's account—which is now incorporated in mine—does.

It satisfies the first desideratum because it indicates a precise mechanism which sits close to the heart of indoctrination, thus telling us when indoctrination occurs

and when it does not. With Yaffe, we can say that indoctrination occurs any time the subject's pattern of response to reasons is being manipulated, but without this sort of manipulation there is no indoctrination.

It satisfies the second desideratum because it is a motive-based characterisation of indoctrination. It reveals an aspect of indoctrination that makes it objectionable to us, namely its manipulative nature. This is an epistemic harm, and it explains why we feel and dread that indoctrination meddles in the cognition of the subjects.

In addition to satisfying the minimum desiderata, Yaffe's account is especially convincing and provides the ideal foundation to develop a fully fledged epistemological account of indoctrination for two main reasons.

First, it allows us to understand indoctrination independently of the notion of education, unlike the tripartite model. By adopting Yaffe's view, we can explain indoctrination not only in the education setting, but outside of it too. We can say that indoctrination occurs *any time* a manipulator shapes a subject's pattern of response to his advantage, not just when a teacher indoctrinates his pupils. For instance, we can explain why political propaganda, religious rituals and some forms of communication—such as biased, partisan news and value-laden, repetitive advertising—can have an indoctrinating effect. By the same token, we can explain better why—even outside the classroom—not every process of acculturation is a case of indoctrination. It is not because it happens in the family, in the office or at the football club, as opposed to school. It is because in these contexts we are not necessarily being manipulated.

Crucially, under Yaffe's account, indoctrination is not necessarily dependent on those factors that are characteristic of the education setting: method, content and intention. Instead, indoctrination essentially depends on whether the subject ends up in the relevant mental state, i.e. a reasoning pattern that serves her manipulator's ends. So, even in the classroom, our intuition that a pupil has been indoctrinated ultimately derives from our judgment about her state of mind as a result of the teaching, not the teaching strategy her educator employed.

The second advantage of incorporating Yaffe's analysis into an epistemologi-

cal account of indoctrination is that it easily lends itself to considerations of epistemological nature. That is due to his focus on the notion of pattern of response to reasons. By highlighting this key notion, Yaffe shows that indoctrination affects how people process information, and how they make decisions as a result of that. So, the next natural step is to dig deeper into this phenomenon, to understand more in detail how it works.

Before moving on, I should address a pressing question: How does Yaffe's account relate to our intuition that indoctrination is somehow closely linked with brainwashing? In popular culture—especially in fiction and in political rhetoric—there is the idea that indoctrination consists in the annihilation, conformation, or control of the subject's thoughts and desires. While this conception of indoctrination—especially when captured by buzzwords such as 'mind control' or 'menticide'—sits somewhere between trope and myth, it could be argued that there are reasonable ways to account for indoctrination which are closer to the concept of brainwashing than to Yaffe's analysis.

One way to do that is to argue that indoctrination is not in the business of reasons, but rather of reflexes. According to this view, when we talk about the way an agent responds to reasons we are not seriously talking about reasoning, but rather of reacting to stimuli. Brought to the extreme, this idea suggests that the subject of indoctrination is somewhat like Pavlov's dog: Whenever the owner rings the bell, the dog expects to be fed, like it was trained to. The subject of indoctrination has no more reasons—in the proper sense of the word—to comply with the manipulator's will than Pavlov's dog has reasons to expect food. If so, it would be more appropriate to say that the subject of manipulation, just like the dog, has been conditioned. More generally, we could say that indoctrination is reducible to conditioning: It is less of an inculcating of beliefs and reasons, and more of an inculcating of behaviour.

Another way to do it is to defend the view that indoctrination damages the cognitive capacities of the subject, whose rationality is thus diminished. This diagnosis of the subject's mental state is meant to explain how she was reduced to intellectual

servitude by the manipulator, and why her reason cannot overpower the influence of the manipulator. An alternative way to express this view is to say that indoctrination inhibits the subject's critical thinking, which explains why supposedly the subject cannot exercise critical thinking to assess her situation and break free from the manipulator.

These views may capture some phenomena in the world. There is no reason to deny the existence of pavlovian conditioning and techniques for numbing a subject's rationality. Nor should we dismiss the possibility of a correlation between these practices and the correct application of the term 'indoctrination'. For instance, the not so rare use of drugs in apocalyptic cults, and forced labour in authoritarian regimes—where cults and regimes are paradigm cases of indoctrination—surely can deteriorate the mental conditions of the subjects to some degree.

However, there seem to be cases of indoctrination—cases that fit with our concept and are close enough to the paradigm cases of indoctrination—that pavlovian conditioning cannot convincingly, fully explain. A necessary feature of pavlovian—i.e. classical—conditioning is intensive repetition. Behaviours and attitudes are instilled through a repetitious process aimed at causing an individual to respond to a precise conditioned stimulus with a precise unconditioned stimulus. But indoctrination practices do not necessarily involve this sort of intensive repetition. Someone may be indoctrinated into an apocalyptic cult through a continually shifting set of practices, behaviours and manipulations, and as a result exhibit attitudes and behaviours that are not necessarily identical or consistent over time. This undermines any account that equates indoctrination to pavlovian conditioning.

Moreover, we cannot fully explain indoctrination even if we consider more complex forms of conditioning—e.g. operant conditioning—characterised by a repetitious process of rewards and penalties. Someone who is indoctrinated into a cult may be subjected to various types of rewards and penalties over time, or none at all, as a result of which her behaviour also varies along significantly. In general, repetitious stimulation does not seem to be a necessary characteristic of indoctrination.

Against pop culture, Yaffe's account shows that the subject of indoctrination

does not lose her mind, and indoctrination has everything to do with *reasons* in the strong sense of the word. He says:

Victims of indoctrination don't, generally, have any less of a rationale for their choices than they had for the choices which they made prior to the influence of the indoctrination. Indoctrination, in the rather technical sense in which I am using the term, affects how agents respond to reasons. In addition, it also tends to provide agents with reasons as well, since it provides them with new desires that they formerly lacked. (Yaffe, 2003, p. 341)

If Yaffe is right, what happens during the process of indoctrination is that the subject exchanges some reasons for others as a result of her pattern of response changing, and acts accordingly.

Further, against the view that indoctrination is about brutally suppressing rationality, Yaffe conveys the idea that indoctrination can operate more subtly. The indoctrinator does not inhibit the subject's rational capacities, so much as *exploit* them. Paradoxically, the indoctrinator may even have an interest in nurturing the subject's intellect to then bend it and benefit from it. A subject is not necessarily less reflective, witty, or smart than before indoctrination. She may be perfectly aware of her reasoning processes, which she monitors and revises. The only thing that she cannot see is the fact that she is carrying out the agenda of someone else.

All in all, the phenomenon that Yaffe captures with his account of indoctrination is more revealing and more explanatory than its typical competitors. So, it is worth picking up and developing further into a fully fledged epistemological account of indoctrination, which I will do in the next chapter.

Chapter 3

Developing the account

3.1 The emergence of a defective pattern

In the previous chapter, I adopted Yaffe's (2003) account of indoctrination as a foundation for developing my epistemological account. According to Yaffe, indoctrination—understood as a form of manipulation—occurs when a manipulator controls the pattern of response to reasons of his subject. When a subject is indoctrinated, she responds to reasons in a way that is advantageous for the manipulator. Whenever the subject tends to respond to reasons differently from how the manipulator desires, the manipulator takes steps to reverse that tendency. In a metaphor: The subject of indoctrination is running on tracks laid for her by the manipulator. Whenever she is about to derail, the manipulator puts her back on track.

In this chapter, I will pick up the reasoning of Yaffe where he left it. Building on top of Yaffe's discoveries, I will try to uncover more facts about indoctrination, and especially facts that constitute the epistemology of indoctrination. To start off, in this section I will speculate how such manipulation technically happens: What exactly it is for, and what the indoctrinator has to do to make it happen. The first questions I need to ask are: How does the manipulator place the subject back on track? What pushes the subject off track in the first place?

One way to answer these questions—which I believe is incomplete—is to say that the manipulator can control a subject's response to reasons by controlling the information she has. By censoring uncomfortable information and isolating the sub-

ject from information sources, the manipulator ensures that she only ever responds to information that is convenient for him.

Under this view, a subject escapes the manipulator's control when she acquires information that is inconvenient for him. What throws her off track is information that undermines the beliefs the manipulator wants her to have. To maintain or take back control over the subject the manipulator needs to prevent this by concealing the undermining information from her.

Information control operates similarly to what Yaffe calls coercion. Through coercion, we give the subject a reason to obtain her compliance, like the armed robber who makes the bank employee hand him the money by threatening her. Similarly, through information control we withhold information from the subject to obtain her compliance; the idea is that the subject cannot act upon reasons she does not know she has. In both cases, we influence the behaviour of the subject by selecting which reasons are available to her. Crucially, to do so we rely on her capacity to respond to reasons in an ordinary, predictable way.

Information control is not the ideal referent for our term 'indoctrination'. First, not all subjects of indoctrination are subjected to information control. Second, keeping information from people is not the most distinctive aspect of indoctrination. Rather, indoctrination is better characterised as an attempt to control people's beliefs and behaviour on the assumption that they will have access to information and evidence that undermines their beliefs. When effective indoctrination is in place, the manipulator does not need to worry about the information flowing to the subject, because the subject will treat all information in a way that is convenient for him anyway. So, to explain indoctrination, we have to assume that the subject is not necessarily isolated from all the evidence that could undermine the beliefs the manipulator wants her to have. Most importantly, she is not necessarily isolated from evidence that can alter her pattern of response to reasons.

Consider the following example. In a totalitarian regime, where the cult of the leader is ingrained through indoctrination, most of the population has bought into it. That is, the average citizen takes whatever the leader says to be a reason to

admire him, to believe him, to support him and to be grateful for his leadership. In this scenario, the average citizen who has been successfully indoctrinated does not take—as they should—the leader’s words as reasons to question whether the nature of his leadership is legitimate, whether it causes more harm than good, and whether the leader’s statements are true. The pattern of response to reasons of the average citizen, then, is convenient for the exponents of the regime, who would do anything to preserve the status quo.

One day a pamphlet starts circulating. It illustrates some of the wrongdoings of the leader, some of his lies, and some of the contradictions in the regime. Surely, this information is not enough to make the average citizen instantly change their mind about the leader. But it is sufficient for triggering a change in their pattern of response to reasons. If the pamphlet produces the effect it was designed for, the average citizen will start taking the words of the leader as reasons to evaluate his leadership, as opposed to adore him, and the cult of the leader in the regime will be altogether undermined.

The pamphlet poses a threat to the regime, in that its contents have the power to change the pattern of response to reasons of the average citizen. So, as we already suspected, the example of the pamphlet shows that it is reasons¹ that push the subject off the tracks the manipulator has laid for them. But more precisely, the derauling is prompted by facts and information that can be taken by the subject as reasons to alter their responsiveness to reasons. So, it is of the utmost priority for the manipulator to prevent the subject from being pushed off track by those reasons, i.e. to take steps to revert the tendency of the subject to respond to reasons differently. How does a manipulator do that?

Let us consider the example again. The most obvious strategy for an authoritarian regime to contain the damage of dissident information is censorship—i.e. information control. However, censorship failed this time around, as the pamphlet has begun circulating. So, the regime must rely on other practices to make sure that *even if* the average citizen reads the pamphlet, their responsiveness to reasons

¹Other factors, such as emotions, may contribute to changing the subject’s pattern of response to reasons. Here, I focus on reasons, and the measures the manipulator takes to counter them.

will not be affected. Being a totalitarian regime, various institutions are set up to address the threat. First of all, the average citizen is surrounded by a community of successfully indoctrinated people ready to explain *that* and *why* the pamphlet is nothing to worry about: the parish, the school, the youth centre, and so on. Second, major propaganda organs and campaigns are there to illustrate *that* and *why* the pamphlet ought to be ignored. Third, in the past citizens have been given reasons to now understand *that* and *why* the pamphlet is none of their concern; therefore, it only makes sense for them to ignore it, should they come across it. Essentially, one way or the other the exponents of the regime counter the reasons of the pamphlet with their own.

This is an example of how manipulators take steps to place the subjects of indoctrination back on track when they derail: by *explaining away* any reasons the subjects might have to alter their pattern of response to reasons. Notice that the subject's derailing can be unplanned, and just be the effect of neutral causal forces. The subject's responsiveness to reasons is not necessarily the object of contention between manipulators and saviours. It is simply someone's cognitive faculty, which the manipulator wants to control.

But there is more. Once someone's faculty for responding to reasons is under the manipulator's control, it becomes defective. As an effect of not being allowed to change, the pattern of response to reasons of the subject starts producing outcomes that are less than ideal—except for the manipulator.

Consider the follow-up of the example. Two siblings, fully indoctrinated into the cult of the leader of the regime, come into possession of the pamphlet. The brother glances sceptically at it, reluctantly opens it, skims through it quickly, and finally gets rid of it. He cannot be bothered by this sort of rubbish. The leader is obviously brilliant, and his deeds are well documented, so whoever says the contrary is either lying or lacks the most basic common sense. By contrast, when the sister puts her hands on the pamphlet, she reads it thoroughly, intellectually stimulated and outraged at the same time. The pamphlet, suggesting that the great leader of the regime is a criminal, is only proof of how great he is. For example, the pamphlet

says that many countrymen feel oppressed. But then—she thinks—these men are the ones who need the regime to discipline them the most! It says the government is struggling to solve some long-time problems. But then imagine the challenge of governing an ungrateful nation! It recounts events in which he committed an abuse of power. Oh, such a strong leadership we are blessed with! When finished reading the pamphlet, which she will treasure in her memory, the sister has an increased conviction that the leader is a hero and a saint.

This example shows two ways in which a subject of successful indoctrination responds to the reasons that they have to change the way they think. Like the brother, one way is to dismiss those reasons before even listening to them. The facts and information that constitute the challenging reasons are ignored because the manipulator provides—or has long since provided—reasons to do that. Like the sister, another way to respond to challenging reasons is for the subject to process them in a way that is convenient for the manipulator, instead of processing them correctly. The manipulator provides the subject with the tools to do it: labels, arguments, motivations. As a result, whatever reasons the subject has to derail are now interpreted as reasons to hold onto the tracks even more firmly. However the manipulator wants the subject to behave, they will do it now with even more conviction than before.

If we *combine* these two types of response to reasons, we obtain a pattern of response that explains away any reason the subject might have to change the way they reason. Every fact and piece of information that the manipulator does not want them to have, they will call it nonsense, misinformation, or disinformation. It is not that the subject is unaware of reasons to change the way they think, it is that they acknowledge them and consciously rule them out.

This is a problem. Systematically explaining away all information that is uncomfortable for the manipulator is a *defective* pattern of response to reasons. In particular, such a pattern is defective in that it violates the general evidentialist principle “that one ought only to base one’s beliefs on relevant evidence (i.e. evidence that bears on the truth of the proposition) that is in one’s possession” (Chignell, 2018). On one hand, when the subject of indoctrination ignores the relevant evidence—like

the brother in the story—they fail to update their beliefs accordingly with the evidence in their possession because they treat the relevant evidence as if it were irrelevant. On the other hand, when the subject of indoctrination twists the significance of the relevant evidence—like the sister—they also violate the evidentialist principle, because they update their beliefs in a way that does not reflect the evidence. Unless the siblings have—which they do not—some higher-order justification to ignore or twist the uncomfortable evidence, then they are wrong in doing so. Their pattern is defective.

To be clear, ‘defective’ does not identify a single pattern of response to reasons. Rather, it is a statement that we can make about the quality of a pattern. Everyone reasons differently, and subjects of indoctrination reason differently from each other too; as seen in the example, the brother responds to the same pamphlet differently from his sister. But there is one thing that the patterns of all subjects of indoctrination have in common: They share the same defect, i.e. they explain away facts and information that are uncomfortable for the manipulator.

If my choice of examples is adequate, and the analysis is correct, then indoctrination—as described by Yaffe—introduces a defect in the reasoning pattern of the subject. The production of this defective pattern is how the manipulator keeps the subject compliant in the face of any challenging input they might receive.

In sum, if indoctrination is a form of manipulation where the pattern of response to reasons of the subject is continuously shaped and corrected by the manipulator, then being indoctrinated is characterised by the constant interfering of the manipulator with the reasons the subject has to change the way she thinks. Such interference is not simply carried out through information control, by simply isolating the subject from information that could undermine the beliefs the manipulator wants her to have. Rather, it is carried out by giving the subject further reasons to process the information in a way that is convenient for the manipulator, so that the potentially undermining information does not end up undermining any beliefs.

By treating the subject in this manner, the manipulator ensures that the subject’s pattern of response to reasons never substantially changes. In doing so, the

manipulator renders the subject's responsiveness to reasons defective, in that the subject systematically explains away all the facts and information that the manipulator does not want her to assimilate. The defective pattern consists in either dismissing them or processing them incorrectly: Instead of being taken as reasons to change one's way of thinking, they are taken as reasons to reinforce it.

We can consider this phenomenon the cognitive mechanism at the core of indoctrination. From an epistemological standpoint, what goes on when indoctrination occurs is the production of a defective pattern of response to reasons in the mind of the subject. By introducing the defect into the subject's responsiveness to reasons, the manipulator ensures that the subject will systematically choose to do his bidding.

In conclusion to this section, I should clarify what I mean when I say that the defective pattern is convenient for the manipulator. The word 'convenient' seems to imply that the manipulator always benefits from instilling the defective pattern in his subject. But that cannot be true, because there are situations in which a manipulator successfully indoctrinates a subject, and yet does not benefit from doing so. In some cases, for instance, indoctrination backfires, i.e. the manipulator ends up indoctrinating his subject into beliefs and behaviours that are inconvenient for him. Further, there are cases of group indoctrination where there is no single manipulator for whom the indoctrination is beneficial. Finally, there are cases in which indoctrination is no longer to the advantage of the manipulator, either because his interests have changed, or he has died. If we were to maintain that indoctrination always benefits the manipulator, then we would also have to explain why these cases do not disprove our claim; but this is not what I maintain. Instead, when I say that the defective pattern is convenient for the manipulator, I mean that it is *conducive* to benefits for him.

When performing indoctrination, the manipulator instils in the subject a pattern of response to reasons that *tends* to produce a response that is desirable for the manipulator. In doing so, the manipulator puts the subject on the right track to support his cause, but crucially he does not coerce the subject to do so. The control

of the manipulator only extends so far: He has direct control over the subject's pattern of reasoning, but not over the outcome of that pattern. For this reason, even if the subject's pattern of response to reasons is convenient for the manipulator, the actual response to reasons may not be.

Let us go over the apparently problematic cases again, starting with the case of backfiring indoctrination. Consider the following example. The leaders of a church indoctrinate their followers into the value of purity by instilling in them a pattern of response to reasons that results in a strive for purity, a strive the church has an interest in supporting. At first, the followers perform the cleansing rituals, learn to entertain pure thoughts, and avoid corrupting influences as desired by the leaders. In time, however, the followers start questioning whether the cleansing rituals are as effective as they could be, they notice that the leaders' words do not come from a place of absolute purity, and begin to stir away from the leaders. Eventually, they leave the church to start their own, reformed and purer. This is not convenient for the original church leaders, who did not want to lose their flock, but it is a case of indoctrination because the leaders successfully put their followers on the path to purity. This is a case of indoctrination backfiring because the very same pattern that was initially beneficial for the leaders ended up motivating the followers to leave the church and starting their own. Backfiring is, simply, one of the possible outcomes of indoctrination.

As this example shows, indoctrination does not guarantee any benefits to the manipulator. Instead, by instilling and tracking the defective pattern, a manipulator ties his interests to the behaviour of a subject; as a result of indoctrination, he has stakes in the subject's response to reasons. Like a gambler, the manipulator *bets* that the defective pattern will cause the subject to act as desired, but that is not a given. Nevertheless, a talented manipulator will have better chances of benefiting from indoctrination than a less talented manipulator.

The case of group indoctrination has a simpler solution: Where no single manipulator benefits from the subject's behaviour, it is the whole group, organisation, or institution the manipulators are members of that should be considered the primary

beneficiary of indoctrination. So, group indoctrination is not a case of indoctrination without beneficiaries. Organisations are entities that can benefit greatly from indoctrinating people, and if an organisation is responsible and benefits from the manipulation performed by its members, then that is a paradigm case of indoctrination.

Finally, there is the case of a manipulator for whom his subject's defective pattern is not convenient anymore. If the defective pattern does not benefit anyone, how can we distinguish indoctrination from other forms of defective thinking that involve the same pattern? Let us consider two cases separately. First, the manipulator whose interests have shifted, and thus the subject's defective pattern is not beneficial to him anymore. The key here is to remember that the only relevant criterion for indoctrination is whether the manipulator instils, monitors and corrects the pattern of response to reasons of the subject so that it tends to serve his agenda. But that is not what happens here. A manipulator who changes his goals, and does not adjust the subject's pattern of reasoning to the new goals, is a manipulator who has let go of his subject. He has either lost grip or renounced his role as a manipulator. So, this is not a case of indoctrination without beneficiaries; rather, it is not a case of indoctrination at all. For this reason, it makes sense to say that the subject has been indoctrinated in the past, but the indoctrination is not ongoing.

Second, the manipulator who dies and thus cannot benefit from the subject's defective pattern. Like the previous case, this case does not satisfy the sole criterion for indoctrination, i.e. that a manipulator instils, monitors and corrects the pattern of response to reasons of the subject. Indeed, it is reasonable to say that the subject has been indoctrinated, but is not currently being indoctrinated. So, this is also not a case of indoctrination, and *a fortiori* not a case of indoctrination without beneficiaries. Simply exhibiting the defective pattern is not equivalent to indoctrination, even if that pattern was once instilled and manipulated by a now-defunct indoctrinator.

The case of the deceased manipulator is an interesting one because it reveals how indoctrination relates to culture in general. The death of the manipulator leads to two possible scenarios. First, a new manipulator or group of manipulators can

pick up the mantle of the deceased manipulator, tie their interests to the behaviour of the subject, and manipulate her pattern of response to reasons so that it becomes conducive to benefits for them. The indoctrination programme, in other words, outlives the original indoctrinator. Second, it can be the case that nobody picks up the mantle, and the subject of indoctrination is left with a defective pattern of reasoning and a set of beliefs, values, and behaviours that the manipulator once wanted her to have. In time, she may or may not derail from them, depending on how deeply ingrained the defective pattern is. But crucially, these elements are now part of her *culture*, which she can transmit, perpetuate, and preserve even though it does not benefit anyone anymore. This strikes as true, when we think of the variety of beliefs, values and behaviours, and even patterns of reasoning, that are now staples of our society, and constitute the legacy of past indoctrinators, who originally promoted them to advance their agendas.

In sum, indoctrination is manipulative in nature and always tends to benefit those who practice it. By instilling the defective pattern and manipulating it, the indoctrinator plants a seed that tends to grow conveniently for him. However, this is just a tendency, so the seed can sometimes grow in a way that is not convenient for the manipulator. This is why indoctrination can sometimes backfire. In any case, there is indoctrination insofar as the manipulator tracks the subject's pattern and has something to gain from it. When the manipulation stops tracking the pattern, out of choice or accident, then we say that the target of the manipulator is no longer a subject of indoctrination. Instead, we say she has been indoctrinated, and the defective pattern and its resulting views are now part of her culture.

3.2 One pattern, different power structures

Up to this point, I have accepted Yaffe's account of indoctrination and argued that what he qualifies as the manipulation of the subject's pattern of response to reasons essentially consists in installing a precise defect in that pattern.

While Yaffe's account, as he presents it, satisfies the two minimum desiderata that I posed for what counts as a convincing account of indoctrination, my elabora-

tion of Yaffe's account may cause some suspicion in this regard. The desideratum inspired by Shiffrin's notion of a motive-based characterisation is satisfied by both accounts, because I do not significantly depart from Yaffe's idea that we have an epistemic harm going on, and more precisely a kind of manipulation. However, as already argued, Yaffe puts his finger on a feature of indoctrination that is exclusive of indoctrination: the manipulation of the pattern of response to reasons. I, on the other hand, by adding the information that indoctrination consists in the production of a defective pattern, may seem to have weakened the strength of Yaffe's original statement.

Indeed, the defective pattern of response to reasons that indoctrination produces is not exclusive of the subjects of indoctrination. The same pattern is also typical of other groups of people, who have not been indoctrinated but have undergone different processes instead. Two processes I have in mind, but there might be more. One is adhering to echo chambers, and the other is acquiring what I call sticky beliefs. In other words, echo chambers and sticky beliefs can produce the same defective pattern as indoctrination does.

Does this mean that, when I further elaborated Yaffe's ideas, I developed an account that does not satisfy the desiderata anymore? No, because I did not altogether reduce indoctrination to the production of the defective pattern. Indoctrination is not defined *only* by such production; other factors are essential and characteristic, especially its manipulative nature. We should not pay any less attention to the manipulation aspect of indoctrination, than to the cognitive mechanism at its heart.

But in the face of two phenomena—echo chambers and sticky beliefs—that produce the same defective pattern of response to reasons, it is important to make explicit what differentiates indoctrination from these two, and everything else. And that is the distribution of power between the people involved in indoctrination, i.e. the manipulator and the subject. Such distribution is what transforms into manipulation the production of the defective pattern. The same distribution of power is not associated with echo chambers and sticky beliefs, which therefore do not translate into manipulation.

To summarise, indoctrination may produce the same defective pattern of response to reasons as other phenomena; nevertheless, indoctrination is a form of manipulation, whereas the other phenomena are not. That is due to the distribution of power among the people involved in each of these scenarios.

In the next subsections (a and b), I will substantiate these claims. I will show in turn that echo chambers and sticky beliefs produce the same defective pattern of response to reasons as indoctrination, but they are not forms of manipulation.

a. Echo chambers

Echo chambers are social epistemic structures that can be found online and offline, though they are typically discussed together with filter bubbles relatedly to social networks. Nguyen (2020)² defines ‘echo chamber’ as follows:

An epistemic community which creates a significant disparity in trust between members and non-members. This disparity is created by excluding non-members through epistemic discrediting, while simultaneously amplifying members’ epistemic credentials. Finally, echo chambers are such that general agreement with some core set of beliefs is a prerequisite for membership, where those core beliefs include beliefs that support that disparity in trust. (p. 146)

Nguyen also reports an example of echo chambers from the literature: the right-wing echo chambers in the USA, and particularly the one surrounding radio commentator Rush Limbaugh. By consistently attacking the mainstream media, developing a private language, and providing counter-explanations for contrary views, Limbaugh completely discredits all other sources of information, exacerbates the insularity of his community of listeners, and undermines not only the views but also the trustworthiness of anyone who expresses contrary views. The result is a community of members who entirely depend on Limbaugh for information, who assign very high levels of trust to each other and deem non-members unreliable, malicious or dishonest.

²Nguyen is but one of a cohort of authors who discuss the topic. These include Goldberg (2011), Jamieson and Cappella (2008), Pariser (2011) and Sunstein (2009).

If this characterisation of echo chambers is correct, then it is easy to show that members of echo chambers and subjects of indoctrination share the same defect in their pattern of response to reasons. Similarly to how the brother ignores the pamphlet in the regime example, in echo chambers information and voices coming from outside are automatically *dismissed* as false or not worth paying attention to whenever they are espousing ideas that conflict with the defining commitments of the echo chamber. Similarly to how the sister devours and twists the words of the pamphlet in the regime example, in echo chambers “members can be brought to hold a set of beliefs such that the existence and expression of contrary beliefs reinforces the original set of beliefs” (Nguyen, 2020, p. 147). This is a phenomenon that Nguyen calls *disagreement-reinforcement mechanism*, and it comes in many forms. One of them is *evidential pre-emption*,³ a mechanism by which “new contrary testimony is neutralized, because it was predicted by past beliefs” (Nguyen, 2020, p. 147). This creates a dangerous feedback mechanism because “by making undermining predictions about contrary testimony, inside authorities not only discredit that contrary testimony, but increase their trustworthiness for future predictions” (Nguyen, 2020, p. 147).

The overall effect of the disagreement-reinforcement mechanism and of the systematic dismissal of outside information is that when echo chamber members have reasons to change the way they think, they do not acknowledge them as such. Instead, they consciously brush them off or think of them as reasons to hold more strongly onto their own beliefs and reasoning processes. This means not only that their beliefs remain unchallenged, but also that their pattern of response to reasons is crystallised—it does not change even though it should. For this reason, their pattern of response to reasons is defective in the same way the pattern of an indoctrinated person is.

As with subjects of indoctrination, we should be careful in automatically declaring all members of echo chambers completely divorced from reason. In the previous chapter, I argued that the cognitive capacities of the subject of indoctri-

³Nguyen credits Endre Begby with first describing this mechanism.

nation are not necessarily inhibited or damaged, but rather exploited by the manipulator, who ideally bends the subject's intellect to his will. Even if the subject of indoctrination always turns out to do the manipulator's bidding, she does not lack reasons to behave the way she does. Similarly, the echo chamber member, who is isolated from most sources of information and finds herself surrounded by people who discredit outsiders, should not be automatically deemed as acting for no reason when she participates in the dynamics of the echo chamber.

Consider *mentalism*, an internalist conception of epistemic justification where “a person's beliefs are justified only by things that are . . . internal to the person's mental life” (Feldman and Conee, 2001, p. 2). Under this view, and depending on which features of someone's mental life we take to be *justifiers* for their beliefs—for instance, coherence and joint consistency with prior beliefs, trusting sources that they consider reliable, etc.—echo chamber members do not appear so detached from rationality anymore. As Nguyen notes:

Once the discrediting beliefs are in place, the ensuing beliefs and actions of the echo chambers' members are surprisingly close to rational. In fact, we can easily imagine alternative scenarios in which a very similar set of beliefs were appropriate and veristic . . . But if such beliefs become implanted in an inappropriate context, they can lead their believers entirely astray. (Nguyen, 2020, p. 149)

But then what is the difference between echo chambers and indoctrination? Is the example of Limbaugh not an example of indoctrination? As anticipated, the difference lies in the way power is distributed in the indoctrination setting and the echo chamber setting.

Indoctrination and echo chambers are two different socio-epistemic environments in terms of power structure. On one hand, there is indoctrination, which is a pyramidal structure, at the top of which is the manipulator, who is the epistemic authority, while at the bottom of the pyramid there are the targets of indoctrination, who are the epistemic subjects. Indoctrination is imposed by the manipulator onto the subjects: It is carefully engineered before execution, so that the ends of

the manipulator become those of the subjects, and the subjects come to reason like the manipulator wants them to. Moreover, the manipulator is responsible for keeping the subjects on track, though intermediaries between manipulator and subject (educators, functionaries, media, etc.) may be involved.

This power structure is evident in most paradigm cases of indoctrination. For example, in the education system, the teacher is the head of the class, and the pupils are subjected to the teacher's authority. Depending on the education system, the teacher has the authority to tell them how to behave in class, what to do in class and at home, what to read, learn, etc. For another example, an authoritarian leader can issue laws to dictate what the citizens are allowed or ought to do, and he has the power—even if self-legitimised—to enforce that.

On the other hand, there are echo chambers, which are essentially communities of peers. Here, every member is responsible for thickening the walls of the echo chamber and keep their peers on track. Orthodoxy is established collectively, and the dissidents are collectively ostracized. To be sure, charismatic personalities may attain more epistemic authority than other members, like Limbaugh in his echo chamber. But having a charismatic leader—or a leader at all—is not an essential feature of echo chambers. There are many examples of echo chambers in which epistemic authority is homogeneously distributed, especially online. For instance, women-hating forums online are hangouts for men who initially simply share some assumptions about women, and later construe together from the ground up a whole subculture based on those assumptions. Nobody indoctrinates them into that subculture; we could say that, if anything, they “indoctrinate each other”. They corroborate each other's beliefs and forge their private language to express them; most importantly, they do not carry out the will of anybody but their own, because they are not being manipulated.

For clarity, let us say that in all cases of indoctrination the manipulator has both power and formal authority over the subjects. He has power in the sense that he exercises some sort of influence, formal or informal, over the subjects: He might have an aura of trustworthiness or wisdom, he might induce fear, or he might appear

legitimised by some holy entity. But most importantly, he has formal authority over them: He is a teacher, a cult leader, a party leader, a policymaker. In echo chambers, a member can become more influential than others because of his charisma; in this sense, a charismatic member like Limbaugh becomes a powerful figure in his community. But even so, there is no formal authority in the echo chamber: someone who is officially in charge of writing the laws or enforcing them. And indeed, Limbaugh does not have this sort of authority.

Importantly, the distinction between indoctrination and echo chambers does not lie in the intentions of the parties involved, nor their maliciousness. We often paint the indoctrinator as an evil mastermind, and consider defective reasoning a side effect of echo chambers. This is a mischaracterisation of both phenomena. On one hand, we should concede that indoctrination can happen without bad intentions: What matters for us to say that indoctrination is going on is that the indoctrinator has control over the subject's responsiveness to reasons. This is the case of good-willed parents. On the other hand, as I will discuss later, the setting up of an echo chamber may be an entirely deliberate, malicious act.

All in all, indoctrination and echo chambers can be identified as different socio-epistemic structures which nevertheless introduce the same defect in the pattern of response to reasons of the participants. The subject of indoctrination exhibits the same defective pattern as the member of the echo chamber, in that they both systematically explain away any reasons they have to change the way they think; the reasons to do so are provided to the subject by her manipulator, and to the member by her peers.

b. Sticky beliefs

In contrast with the established literature on echo chambers, *sticky beliefs* is a new coinage. With it, I would like to group those beliefs that are characterized by the capacity to produce a defective pattern of response to reasons, analogously to indoctrination and echo chambers. A belief is sticky when, once acquired, it produces the defective pattern.

When the believer acquires a sticky belief, she faces a significant barrier in

dropping it in the face of counterevidence, as well as a barrier in overcoming the defective pattern altogether. Hence the term ‘sticky’: It conveys the idea that once the belief is installed, it is quite difficult to get rid of.

Notice that I call ‘belief’ the propositional content that a person may come to believe, which may be either atomic or a conjunction of propositions. To clarify what I am talking about, let us consider, in turn, two types of sticky beliefs: conspiracy theories and just world beliefs.

Conspiracy theories are a prime example of sticky beliefs. Consider the most basic form of conspiracy theory: the idea that there is a conspiracy. This belief is sticky because once a person believes that there is a conspiracy, it is very difficult to persuade her otherwise. That is because the very concept of conspiracy provides various explanations as to why there may seem to be a lack of evidence or even counterevidence for it (from now on, I refer to both as *uncomfortable evidence*). Such explanations for uncomfortable evidence include cover-ups, disinformation, and the appeal to secret evidence, to name a few. While these explanations can be invoked to explain away many kinds of statement, they particularly make sense under the assumption that there is a conspiracy, because conspiring involves these sorts of activities by definition.

A person who believes that there is a conspiracy is naturally inclined to explain away any accusation that her belief is under-supported by evidence or contradicted by it, by invoking cover-ups, disinformation, etc. Like the subject of indoctrination and the member of an echo chamber, the person has two types of response available to counter uncomfortable evidence. She can dismiss it by simply labelling it as misinformation and brushing it off her back. Alternatively, she can twist it in a way that supports her belief in a conspiracy, instead of undermining it: She labels it as misinformation, and then she takes the alleged misinformation as further proof that there is a conspiracy. Additionally, there is a sort of vicious feedback mechanism in place, by which the person, thinking that the status of her conspiracy theory is not-disproven—while de facto it is only non-disprovable in her eyes—acquires increasingly more confidence in that theory.

Dentith (2019) registers something that supports this analysis. He addresses the fact that conspiracy theories are often depicted as beliefs supported by suspicious evidence. Against this view, he argues that “there is no *prima facie* justification for a suspicion of the kinds of evidence conspiracy theorists are alleged to rely upon” (Dentith, 2019, p. 2244). Rather, “if there is an issue with the evidence used in support of conspiracy theories, then it is an issue of principle: the evidence is being abused or just not being used appropriately” (Dentith, 2019, p. 2244). He defends this view by convincingly showing that the type of evidence employed in conspiracy theories and often considered suspicious is not exclusive of conspiracy theories and not necessarily problematic outside them.

Dentith essentially notes that the conspiracy theorist actively acknowledges the relevant information—the one that should instil some doubt about the truth of the conspiracy theory—and nevertheless she processes that information in a dysfunctional way. That, I believe, is because the pattern of reasoning of the believer is defective, and such defect is caused by the very nature of conspiratorial beliefs: They are sticky, i.e. they lend themselves to systematically explaining away uncomfortable evidence.

Although the conspiracy theorist and the subject of indoctrination share the same defective pattern, there is a crucial difference between the two, which is the reason why we say that the conspiracy theorist is not indoctrinated. Once again, to see this we ought to look at the socio-epistemic structure of the two situations.

On one hand, indoctrination is a form of manipulation, where a person who has influence and formal authority over the other manipulates his subject. So, there are at least two persons involved and a hierarchy. On the other hand, when a conspiracy theory takes root in a person’s mind, no more than one person needs to be involved, the believer, and she may have acquired the belief by various means, e.g. she may have heard of it, made it up or been persuaded of it. Crucially, the social context in which the conspiracy theorist first acquires her belief is irrelevant: For the defective pattern to arise, the mere belief into that particular propositional content is sufficient.

The bottom line is that the very propositional content of a conspiracy theory can do what a manipulator by definition does in indoctrination: altering the pattern of reasoning of a subject in a specific defective way. Conspiracy theories produce the defective pattern just in virtue of their content, whereas indoctrination has to be carried out by a person who has influence and formal authority over the subject.

But conspiracy theories are not the only type of belief that has this effect, i.e. producing a defective pattern in virtue of its propositional content. This is the peculiarity of sticky beliefs. Another type of sticky belief are just world beliefs.

Just world beliefs, JWB for short, are a widely studied topic, with ample literature especially in Psychology, Philosophy and the social sciences.⁴ Kasperbauer (2015) characterises them as follows:

The basic idea behind JWB is that people operate under the assumption that the world is a just place, commonly expressed in the psychological literature as “people get what they deserve and deserve what they get.” JWB are considered to be a real psychological phenomenon shared to some degree by all people. (p. 210)

JWB have to do with our understanding of the fundamental nature of the world and represent a cognitive bias which operates in the subconscious. But a most important aspect of JWB is that they are continuously under threat. That is, JWB are continuously challenged by events that constitute reasons to doubt JWB, and even drop them.

Research has consistently found that severe, prolonged suffering, and success owing to luck, are two of the main threats to JWB. Research also shows that we humans do our best—or our worst—to resist threats to JWB, because JWB are low-key comforting and give us the confidence to face the risks that come with navigating the world.

For instance, humans typically respond to severe prolonged suffering with

⁴As Kasperbauer (2015) reports: “The collection of ideas that eventually came to be known as Just World Theory first originated in research conducted by Melvin Lerner and his colleagues” (p. 25) in the mid-Sixties. There has been a rich debate since, which includes contributions by Hafer and Bègue (2005), Hirschberger (2006) and Wakslak et al. (2007).

victim-blaming and respond to cases of success owing to luck by either denying the luck or assuming more effort was put forth even though there was clearly none. As Kasperbauer puts it: “A troubling result of Just World Beliefs [...] is that people tend to dismiss or explain away any threats to their belief that the world is fundamentally just” (Kasperbauer, 2015, p. 217).

Precise accounts of why and how we dismiss and explain away the threats to JWB—i.e. the *reasons* to drop JWB—are found in the literature. I shall not elaborate further here, because it is already clear why JWB are sticky beliefs. When our JWB are challenged, we look away: We dismiss the threats as not relevant, we do not take them to be reasons to revise our JWB. Alternatively, we take the facts that are threatening to be reasons to find further information that will explain away the threat. The harder the challenges, the more stubbornly we turn away, and the more eagerly we look for alternative explanations.

This is due to the vital function of JWB, the attractiveness of a world seen through their lens; which all boils down to the content of these beliefs. Like conspiracy theories and like indoctrination, JWB introduce a defect in the pattern of response to reasons of those who hold them. Like conspiracy theories, this is due to the content of the beliefs; unlike indoctrination, they do not represent a form of manipulation.

3.3 A delicate balance

Up to this point, three main claims have been made. First—this is Yaffe’s claim—indoctrination is a form of manipulation, where the indoctrinator manipulates the subject’s pattern of response to reasons so that the subject will always choose to do his bidding. Second, the manipulation consists in giving the subject reasons to dismiss or explain away any reasons she might have to change the way she thinks, which is convenient for the manipulator. In doing so, the pattern of response to reasons of the subject remains substantially unaltered in the face of reasons to change, and as such is defective. So, we can say that indoctrination produces this defective pattern. Third, the same defective pattern can be produced by other phenomena that

are not indoctrination: Members of echo chambers develop the same pattern, and so do people who maintain certain beliefs that I call sticky in virtue of their particular propositional content.

In the light of these three points, we need to find the delicate balance between *reducing* indoctrination, echo chambers and sticky beliefs to one another, and treating them as *entirely separate* phenomena. That is, the relation between these three phenomena is a complex one, and the challenge is to explicate how they relate to each other and the defective pattern.

First, we should avoid reducing indoctrination, echo chamber and sticky beliefs to one another. As argued in the previous section, the defective pattern emerges under different circumstances, and each case counts as an instance of indoctrination, echo chamber or sticky belief mostly depending on how the power is distributed in that situation. This means we should not think of the emergence of the defective pattern as occurring *despite* the differences in power structure across situations. Rather, it is the power structure that *determines* whether a situation is a case of indoctrination, echo chamber or sticky belief. The difference between indoctrination, echo chambers and sticky beliefs is *structural*. Therefore, it is a mischaracterisation to reduce these three phenomena to one another.

Not only is there is value in attending to our concepts, but it also has practical advantages. For example, we often tend to lump the indoctrinated and conspiracy theorists together, because of the similar way in which they stubbornly, passionately defend their beliefs. But suppose there is a person who has done an extensive job at formulating a conspiracy theory. Not someone who passively accepts it. Rather, someone who read history books, exercised her critical thinking, listened attentively to various opinions and concluded that there is a conspiracy. Further, suppose that her conspiracy theory is wrong and we want to help the person realise it. Attempts at undoing her supposed indoctrination would be a waste of time because she is not indoctrinated.

For instance, it would be useless for us to try and generate *Indoctrination Anxiety* in her. Generating Indoctrination Anxiety is a type of *Etiological Challenge*

which is specifically targeted at indoctrinated people. Not only it generically aims at urging someone to revise their beliefs “by pointing out the contingent causal origins of those beliefs” (DiPaolo and Simpson, 2016, p. 1), like any Etiological Challenge. Rather, it aims precisely at instilling the suspicion that someone’s beliefs are “a product of a method of belief-transmission that’s geared towards making people believe things which they would otherwise reject” (DiPaolo and Simpson, 2016, p. 5).

Our conspiracy theorist, who conducted her research without any external pressure, is not susceptible to this kind of anxiety. Instead, she might be struck by other forms of Etiological Challenge which address the stickiness of her beliefs instead. Keeping this in mind, it seems convenient to accept a conceptual distinction between sticky beliefs and indoctrination, as well as echo chambers.

Second, we should not treat these phenomena as completely separated either. While they differ in their socio-political aspects, we must conceive them as being characterised at heart by the same cognitive mechanism, i.e. the emergence of the defective pattern. Again, acknowledging this is not just for conceptual clarity, it also serves a practical purpose: understanding why and how these phenomena intersect and interact in the world, and dangerously so.

For example, the fact that indoctrination and echo chambers produce the same cognitive effects means that indoctrinators can benefit from creating an echo chamber around themselves. As Nguyen points out, “echo chambers are excellent tools to maintain, reinforce, and expand power through epistemic control. Thus, it is likely (though not necessary) that echo chambers are set up intentionally, or at least maintained, for this functionality” (Nguyen, 2020, p. 149). When an indoctrinator utilises an echo chamber in his favour, he doubles down the exploitation of that cognitive mechanism that indoctrination and echo chambers share.

For a similar reason, indoctrinators can and often use sticky beliefs to consolidate the defective pattern of their subjects. For instance, totalitarian regimes typically spread conspiracy theories to defame their political enemies. Because conspiracy theories are sticky, if a regime manages to tie its consensus to the citizens’

belief in those theories, then it will benefit from how difficult it is to undermine a conspiracy theory.

Conspiracy theories interplay with echo chambers too, but in a different manner. Echo chambers benefit from incorporating conspiracy theories in their orthodoxy more than any other type of theory, because they are extraordinarily effective in dismissing and explaining away outside testimony, thus reinforcing the defective pattern. For example, conspiracy theories are responsible for *inverted corroborative bootstrapping*, a phenomenon where echo chambers “are attributing a problematic form of non-independence to outsiders who are actually independent, and thereby underweighting outside testimony” (Nguyen, 2020, p. 148).

Alongside conspiracy theories, the stickiness of JWB (just world beliefs) can also be piggybacked by an indoctrinator to reinforce the defective pattern of his subjects. The idea that JWB influence our political choices is not new: Kasperbauer argues that JWB constitute a psychological barrier to egalitarianism, essentially because whatever social injustice egalitarianism purports to correct, JWB tell us that that is simply how things should be. It is clear then what power an indoctrinator can gain from leveraging on JWB. A dictator, for example, can match the intuition that the world is a just place with the notion of divine providence. From there, he can claim that divine plans are not for everyone to understand, but he is their designated interpreter. Finally, by cultivating the citizens’ confidence in JWB, he can justify all his actions as part of the nation’s destiny.

These are just some examples of how indoctrination, echo chambers and sticky beliefs easily interact and intersect due to their shared cognitive mechanism, suggesting that—on the epistemological level—these phenomena are not completely separate.

Having illustrated the complexity of the relation between indoctrination, echo chambers and sticky beliefs, here is how I propose we understand indoctrination.

For an analogy, consider the feeling of loss. This is a state of mind characterised by a feeling of abandonment, helplessness, etc. We feel loss on different occasions, for example when we mourn a loved one, when we experience certain

forms of art such as a beautiful tragedy, or when our relationships end. Mourning, aesthetic appreciation and break-up are complex phenomena distinct from each other in many ways, but on a psychological level they have an element in common, the feeling of loss.

Similarly, the defective pattern is a general, purely cognitive phenomenon that manifests itself in various ways. Indoctrination, echo chambers and sticky beliefs are some of its manifestations. They are more complex socio-epistemic phenomena, different from each other in many ways, that nevertheless share a same epistemological component. Although they are distinct from each other on some level, from a strictly epistemological point of view they all boil down to the same purely cognitive phenomenon, the emergence of the defective pattern.

Indoctrination, echo chambers and sticky beliefs are then a family of phenomena that belong to the same kind. Indoctrination is not a kind of its own; rather, it belongs to a kind. What makes indoctrination belong to that kind is the fact that it is characterised by the emergence of the defective pattern. Echo chambers and sticky beliefs belong to the same kind for the same reason.

From an epistemological point of view, the emergence of the defective pattern is a basic cognitive mechanism which articulates itself in increasingly higher levels of complexity: It develops and differentiates into indoctrination, echo chambers and sticky beliefs on a superior level, and into different types of indoctrination on an even higher level. That is how indoctrination in the education, the political, and the religious context boil down to the same deeper phenomenon, namely indoctrination; and in turn how indoctrination, echo chambers and sticky beliefs boil down to the same deeper phenomenon, namely the emergence of the defective pattern.

In sum, granted that these three socio-epistemic phenomena are not reducible to one another, we can say that they nevertheless belong to the same kind. They belong to the same kind because they are expressions—so to speak—of the same mental state, namely the defective pattern, the declension of which occurs in diverse socio-epistemic structures and contexts.

3.4 Revising our worries about indoctrination

The idea that indoctrination belongs to a family of socio-epistemic phenomena characterised by the emergence of the defective pattern compels us to revise the conceptual worries of the tradition.

The main concern in the Analytic tradition is how to distinguish indoctrination from proper education. This concern stems from the assumption that indoctrination is a form of teaching, just like proper education and other forms of bad education. But in the light of the membership of indoctrination to its kind, to call indoctrination a form of teaching now appears reductive and somewhat misleading. If we are open to the idea that indoctrination is a form of manipulation, we find ample evidence in the world and in the literature that indoctrination is not a type of teaching, and rather that teaching is one of the many ways in which someone can be indoctrinated.

So, philosophers in the tradition are right about something: Indoctrination is hard to account for because it is similar to other phenomena. But they are posing the wrong question. Instead of asking how to tell indoctrination apart from proper education, we should be asking how to distinguish indoctrination from other forms of manipulation, and more precisely other forms of flawed reasoning involving the defective pattern, such as echo chambers and sticky beliefs. I have already sketched an important way in which these phenomena differ—i.e. in their socio-power related aspects—and I will say more in the next section, but I suggest that there is plenty of room for further research on the topic.

Another worry stemming from the Analytic tradition—one that however is never explicitly discussed—is how to explain the relation between indoctrination and the other ways in which someone can be indoctrinated, but which do not belong to the education context. We see this worry transpire, for example, through the words of Green (2010). In a passage, he claims that “that propaganda, lies, threats, and intimidation have been used as methods of education is not doubted” (Green, 2010, p. 23). What Green is doing is contraposing the concept of indoctrination to the notions of propaganda and lying, while subsuming everything under the concept of teaching.

This is a strange move: Instead of admitting that indoctrination is not a form of teaching, he prefers to say that propaganda and lying are. But after reframing indoctrination as manipulation, we can safely admit that indoctrination is not a form of teaching, and that propaganda and lying are neither. Further, we can establish a more convincing relation between them: Propaganda and lying are some of the ways in which someone can be indoctrinated, as seen previously in the David Koresh and authoritarian regime examples. So, by admitting that indoctrination is not a form of teaching and that there are other ways to indoctrinate alongside teaching—including practices that do not belong in the classroom—we ease this particular worry of the tradition without raising any new ones.

Finally, a major concern in the tradition are the children. Since indoctrination is conceived as a form of teaching—more often than not schooling—the tradition ignores the question of adult indoctrination. Most accounts share the view that indoctrination can only happen in a short window of time, while the subject is still ignorant and docile—i.e. a child—and thus exclude the possibility of adult indoctrination because adults are assumed to have developed sufficient critical thinking skills to resist indoctrination.

But as we reframe indoctrination as manipulation, it becomes evident that we should talk about adult indoctrination too. All the paradigm cases presented until now involve adults as much as children—the authoritarian regime, the apocalyptic sect, etc. That is because indoctrination is not a deed performed in one go while the child is impressionable. Rather, it is an ongoing manipulation that can continue throughout adulthood, or begin in adulthood. The specific harm of indoctrination is not the exploitation of the child's little capacity for critical resistance, but rather the exploitation of the subject's capacity to respond to reasons, whether infant or adult. The worry, then, should not just be to prevent children from growing up into badly-educated adults. Rather, it should also be how to prevent anyone from being manipulated.

In sum, as we revise our concept of indoctrination, we revise our worries about it. Under my account, some of the traditional worries are completely dissipated,

and others are reframed.

3.5 Exploiting a vulnerability of the mind

My descriptive analysis so far illuminates the fact that indoctrination sits in a complex space between the inculcation of beliefs and the controlling of people's wills. The indoctrinator does not merely drill beliefs that are false or not known to be true. Rather, he manipulates the subject's will in respect of her epistemic practices: He influences her conscious decisions as to which evidence she pays attention to. In doing so, he ensures that the subject only engages with evidence that serves his purposes. Crucially, he ensures that she does not engage with evidence that can lead her to alter her epistemic practices that serve him well.

But in turn, this view illuminates a further aspect of indoctrination. When the indoctrinator manipulates the subject's will with respect to her epistemic practices, he is effectively exploiting an epistemic-cum-cognitive vulnerability of her mind, which allows for the establishment and taking root of the defective pattern.

The defective pattern is a mental state in which we can end up. It does not only emerge as a result of indoctrination. It also emerges spontaneously and uncalled for, as in echo chambers and sticky beliefs. Further, it has a distinctly self-preserving and self-reinforcing nature, in that—once it has emerged—it naturally leads the person affected by it to ignore any evidence that could change her epistemic practices.

Our mind is vulnerable, in that it is subjected to the possible emergence of this pattern—spontaneous and non-spontaneous—and once the pattern has firmly established itself, the mind tends to remain affected by it. Indoctrination is the exploitation of this vulnerability, in that the indoctrinator actively seeks to instil the defective pattern in the subject's mind and intervenes any time in which—against all odds—the subject tends to derail from it.

This, I take it, is ultimately the distinctive mark of indoctrination, what sets it apart from echo chambers and sticky beliefs. In indoctrination, the emergence of the defective pattern is actively pursued by an indoctrinator who exploits his victim's vulnerability, and the defective pattern is actively nurtured by the manipulator once

it has taken root. By contrast, in echo chambers and sticky beliefs there is no pursuit or nurturing of this sort.

This is not to say that we should go back to tying indoctrination to an intention criterion. The indoctrinator does not necessarily tell himself that what he is doing is exploiting a vulnerability which results in the emergence of the defective pattern. In some cases, the indoctrinator only intends for his subject to make the choices he thinks are right, and sees himself as informing the subject of what her best options are; in other cases, the indoctrinator knows exactly what vulnerability he is exploiting. Either way, indoctrination happens, because what the manipulator wants is for the subject to choose to do his bidding, and he takes steps to let it happen. In doing so, he effectively ends up exploiting her vulnerability, by causing the defective pattern to emerge where it would not have, had he not intervened.

So, the relevant and sole criterion of indoctrination is whether the vulnerability of the subject's mind is being exploited by an indoctrinator, who manipulates her will with respect to her epistemic practices. This is the trademark of indoctrination, what distinguishes it from echo chambers and sticky beliefs, as well as everything else.

Adopting this criterion compels us to revise how we conceive indoctrination: It invites us to abandon the popular view of indoctrination as an epistemic misdeed, and rather understand indoctrination as one way in which our mind can fall in the cognitive trap that is the defective pattern. Indeed, we are used to thinking of indoctrination as the *performance* of the indoctrinator. That is, either the bad teaching of a professional indoctrinator who brainwashes the youth, or the bad parenting of adults that have their children indoctrinated as part of their upbringing. But knowing that indoctrination is the exploitation of a pre-existing epistemic-cum-cognitive vulnerability, present in both children and adults, it becomes more appropriate to focus on the *subject's mind* instead.

This is especially true if we are interested in the normative implications of how we conceptualise indoctrination. Typically, in the normative discourse, we are focused on assigning culpability: blaming the evil manipulator, judging the

parents' choices, and protecting the children from acts of human wickedness. But if we take seriously the role of our vulnerability in indoctrination, then it becomes of primary importance to understand how to move about it. We acquire an interest in learning how to protect ourselves from it; we become concerned with preventing the formation of the defective pattern of response to reasons. Further, being aware of our vulnerability—that incidentally is exposed to echo chambers and sticky beliefs too, which raises the stakes—compels us to learn how to recognise and dislodge the defective pattern in ourselves and others, when needed. So, naturally, from this perspective, the task of assigning culpability rightfully sinks in the background. Suggesting that these concerns should be addressed in further research, in the next chapter I will address on one of them.

In conclusion to this chapter, I claim to have identified the epistemic phenomenon which is closest to what we take ourselves to be referring to when we talk about indoctrination and which best explicates its paradigm cases. That is, the exploitation of a precise epistemic-cum-cognitive vulnerability, i.e. our tendency to acquire the defective pattern and remain fixed in it. In keeping with the principles of descriptive analysis, I take it that this discovery—obtained by analysing the paradigm cases of indoctrination and resorting to the literature on manipulation—explicates the kind indoctrination belongs to: A kind that sits in a complex space where the epistemological dimension and the sociological one intersect.

Chapter 4

Ideas for counteracting indoctrination

4.1 Dislodging the defective pattern

In the previous chapters, I offered an epistemological account of indoctrination. With Yaffe, I argued that indoctrination is best understood as a form of manipulation and precisely the manipulation of the subject's pattern of response to reasons. Further, I argued that this manipulation consists in rendering the subject's pattern defective in a precise way: making it systematically dismiss or explain away any reason the subject may have to change their pattern of response, thus escaping the manipulator's control and not choosing to do the manipulator's bidding anymore.

I then argued that this defective pattern is not exclusive of indoctrination, but it also pertains to other socio-epistemic phenomena, such as echo chambers and sticky beliefs. This, I claimed, is why we can consider indoctrination not an epistemic phenomenon of its own, but the member of a family of related socio-epistemic phenomena, which are manifestations of a more general epistemic one, the emergence of the defective pattern.

This reveals that, effectively, what indoctrination does is exploit an epistemological-cum-cognitive vulnerability of the mind. That is, its exposure of the mind to the risk of becoming crystallised in the defective pattern, which can happen in multiple ways. While the defective pattern can spontaneously emerge and take root, as in echo chambers and with sticky beliefs, indoctrination actively

seeks to instil it.

Up to this point, my epistemological account was developed entirely on the descriptive side of the debate. Nevertheless, in keeping with Shiffrin's methodology, it was developed as a motive-based characterisation. So, it incorporates the description of those aspects of indoctrination that make it morally objectionable to us, and points towards the direction we should go in, should we move onto the normative side of the debate. Particularly, it makes it almost inevitable for us—if we accept the account—to try and oppose indoctrination. We have identified a precise epistemic harm that it necessarily produces, so it would be bizarre to defend indoctrination or treat it as unproblematic at this point.

Further, I suggested that the next natural step into the normative side of the debate would be to address the issue of our vulnerability to the defective pattern: what to do about it, how to prevent its exploitation, etc. In this chapter, I will give some ideas on how to counteract indoctrination, i.e. what we can do once the defective pattern has been established and the vulnerability is already being exploited. More precisely, I will focus on finding ways to dislodge the defective pattern—to eliminate the defect.

The ideal way to do this is to find a solution that is feasible even in a scenario where the manipulator has already been removed from the picture. Indeed, one might be tempted to try and dislodge the defective pattern by physically removing the subject from the sphere of action of the manipulator, so that he cannot put her back on track when she naturally tends to derail. But as we have seen, the subject who is deeply indoctrinated will put herself back on track, as a result of her defective pattern. Also, she might not ever elicit the tendency to derail.

This means it is essential to focus on the epistemological dimension of indoctrination, as opposed to just its social dimension: not solely removing the subject from the power-structure typical of indoctrination, but crucially engaging with her on an epistemological level. Understanding how to do this is the task I take on in this chapter, where I assume the manipulator has already been removed.

There is a principal obstacle to dislodging the defective pattern, overcoming

which is the key to start undoing indoctrination; the obstacle is that what seems to be the most reasonable way to help the subject of indoctrination is instead most ineffective and even counterproductive. When helping the subject, the desired outcome is for her to question her epistemic practices, as well as the beliefs that she maintains because of those practices. To obtain this outcome, we are naturally inclined to show the subject evidence that she is mistaken and persuade her to revise her beliefs. However, this resolution does not deliver the results we imagine. Instead of dislodging the pattern, giving the subject evidence that should help her do so only reinforces the pattern. We fail to undo indoctrination by reasoning with the subject because we generate belief polarisation.

4.2 Belief polarisation

For decades, researchers in Psychology have been aware of the phenomenon of belief polarisation.¹ In a nutshell, the empirical evidence shows that when two people, who disagree about a matter of fact, are exposed to mixed evidence bearing on the matter, they tend to hold their initial position more firmly. A priori, this is an unexpected outcome. As Kelly (2008) observes:

It is natural to expect—and perhaps, also natural to hope—that mutual exposure to common evidence will tend to lessen or mitigate our disagreement . . . At the very least, one would expect that exposure to common evidence would not increase the extent of our disagreement.
(p. 612)

And yet, the phenomenon exists and is well-confirmed.

Normative considerations aside (for now), various attempts have been made to describe the mechanism which causes—or consists in—belief polarisation, i.e. to answer a set of “descriptive, psychological questions about how exactly You and I are responding to our evidence so as to generate the relevant phenomenon” (Kelly, 2008, p. 612).

¹For instance, an important study on the subject, but not the first, was conducted by Charles Lord, Lee Ross and Mark Lepper in 1979.

One explanation is elaborated by Kelly, who suggests the following model. He notes two key facts that are supported by empirical evidence. First:

All else being equal, individuals tend to be more adept at detecting fallacies when the fallacy occurs in an argument for a conclusion which they disbelieve, than when the same fallacy occurs in an argument for a conclusion which they believe. (Kelly, 2008, pp. 617–618)

Second, we have a “tendency to devote more thought to evidence which seems to tell against our beliefs than to evidence which seems to tell in their favour” (Kelly, 2008, p. 618), because we tend to devote more cognitive resources to looking for alternative explanations for data that does not match our worldview, than for data that does. So, “all else being equal, the more cognitive resources one devotes to the task of searching for alternative explanations, the more likely one is hit upon such an explanation, if in fact there is an alternative to be found” (Kelly, 2008, p. 618).

Essentially, belief polarisation happens, according to Kelly, due to each person employing double standards for evidence that supports their previously held beliefs, and evidence that does not. On one hand, we have higher standards of scrutiny of unwelcome evidence, so to speak, than for more welcome one. On the other hand, we spend more time looking for alternative hypotheses to explain unwelcome data, than for welcome data. Because of this, we “continue to respond to incoming evidence in the light of our prior beliefs”, and thus “the net effect is that we are pushed further and further apart” (Kelly, 2008, p. 620).

Kelly presents his model in opposition to another one, which he calls *Kripkean dogmatism* and which is inspired by Saul Kripke’s *Dogmatism Paradox*. Kripkean dogmatism, as presented by Kelly, thus represents a second way of explaining how belief polarisation works. According to this model, belief polarisation is generated when someone reasons as follows. When I believe proposition *p*, and I am invested in truth, it seems sensible to ignore any misleading evidence about *p*. Because I know *p* to be true, it follows that any evidence against *p* must be misleading evidence. So, when I am exposed to mixed evidence about *p*, I take the evidence for *p* to be reliable, and the evidence against *p* as misleading. By only taking the

supporting evidence seriously, my confidence in the belief that *p* is reinforced.

Kripkean dogmatism is presented by Kelly as unreasonable behaviour, which makes for an unflattering explanation of belief polarisation. By contrast, Kelly presents his model as capturing a distorting or biasing factor in our reasoning that does not necessarily translate into unreasonable behaviour. While Kelly provides a detailed account of what exactly makes Kripkean dogmatism senseless,² thus implying that it seems unlikely that it constitutes ordinary behaviour, he defends his model for belief polarisation based on empirical evidence.³

There are other ways to account for belief polarisation, including some that are radically different from Kelly's. For instance, while Kelly assumes that belief polarisation is a distinctly epistemological phenomenon, other authors emphasise its sociological dimension. They talk about the influence of peer pressure, tribalism, the fear of appearing indecisive or of admitting being wrong, and so on. So, what are we to make of this proliferation of explanations for belief polarisation? I suggest that we do not need to pick just one.

Although Kelly presents his model as an alternative to Kripkean dogmatism to explain belief polarisation, both Kelly and Kripke may describe existing phenomena. Surely, if the empirical evidence suggests that people behave as Kelly describes, we should accept Kelly's account as true. But even so, we should not automatically exclude the existence of Kripkean dogmatism.

²Kelly argues that Kripkean dogmatism is intrinsically irrational behaviour because the dogmatist does not respect The Commutativity of Evidence Principle, i.e. the fact that "the order in which pieces of evidence are acquired makes no difference at all to what is reasonable for one to believe" (Kelly, 2008, p. 616). Against this principle, the order in which the dogmatist acquires information heavily shapes what she believes and how strongly she believes it. That is because once she starts believing something, all information she is exposed to in a later moment is automatically labelled reliable or misleading in the light of what she already believes.

³It is worth mentioning that Kripkean dogmatism is a variation of Kripke's original Dogmatist Paradox. Kelly's interpretation of the paradox portrays the dogmatist as irredeemably irrational. In his original paradox, however, Kripke poses a clause that disappears in Kripkean dogmatism: The paradox discusses "a subject *S* who genuinely knows that *p*" (Kripke, 2011, p. 22). The premise to his paradox, is that "we are dealing with a subject who really knows. We are arguing that such a subject ought to maintain the dogmatic attitude because any counterevidence really is misleading" (Kripke, 2011, p. 22). By adding this clause to Kripkean dogmatism, belief polarisation does not appear so vicious anymore. For this reason, Kelly made sure to rule out the clause, when proposing Kripkean dogmatism as a possible explanation for belief polarisation. In doing so, he is supported by the sceptical intuition that a subject *S* can never know that they "genuinely know *p*".

For all we know, people could be more or less biased in the way Kelly describes, and those who are less biased may tend to be dogmatists. Moreover, some people may be able to compartmentalise: when it comes to certain topics they elicit Kripkean dogmatism, whereas they are simply biased in all other circumstances. So, it is not absurd to think that Kripkean dogmatism and Kellyan reasoning may co-exist.

Given that there are multiple competing explanations for belief polarisation, we might think that we must pick the most plausible one. But the variety of convincing explanations for belief polarisation suggests that there is a surface-level phenomenon of belief polarisation—the manifestations of which are broadly similar, namely, they involve people with similar evidence adopting more polarised views—while underneath this there are (or could be) a plurality of different phenomena: Kripkean dogmatism, Kellyan reasoning, etc. In other words, we are lumping various phenomena under the label of ‘belief polarisation’.

To see that, consider Hannon’s (forthcoming) idea that political disagreement is nowadays less deep and extensive than the media and the public make it. He claims that “voters are increasingly polarized in terms of their *attitudes* towards each other, even though there has been comparatively little polarization on the *issues*” (Hannon, forthcoming, p. 2). We perceive the disagreement to be wider than it is because often people respond to surveys through what Hannon calls “expressive responding”, that is an inaccurate representation of their beliefs and a caricatural representation of their attitudes towards them. He says:

People engage in this behaviour for at least two reasons: either partisans know the truth but prefer to “cheerlead” when there is nothing to gain from accuracy, or they are ignorant on the issue and they offer a congenial answer as their best guess. Either way, survey responses are not entirely sincere. (Hannon, forthcoming, p. 2)

Assuming that Hannon’s idea is plausible, we have an example of a phenomenon that is not, strictly speaking, an instance of belief polarisation, in that it does not involve the polarisation of *beliefs*—rather of attitudes—but it looks like

it, in that it involves people with similar evidence adopting more polarised views and defending their beliefs with increasingly more confidence.

Views such as this one have the potential of undermining the very notion of belief polarisation. Like Kelly's and Kripke's, Hannon's view explains what might be going on in at least some cases of what we call belief polarisation, but suggests that we are mistaking partisan cheerleading for real belief polarisation.

While we do not want explanations such as partisan cheerleading to do all the work in explaining belief polarisation, we also do not want to discard them, because they are plausible and useful. So, it is in our best interest to accept that phenomena such as the ones described by Hannon, Kelly and Kripke can co-exist.

In fact, Hannon's and Kelly's accounts fit well together: It is possible for two people, who are genuinely becoming polarised on some issue, to also be drawn—on top of that—to express their attitude towards the other faction. Hannon's idea is that what people express does not always match what they believe: What they say is often exaggerated and untrue if there is no incentive for accuracy. But even if this is the case, it does not mean that the polarisation does not happen beneath the surface—only that we need to re-think strategies to detect it, for instance by providing an incentive for accuracy.

So, I will assume the view that there is a surface-level phenomenon we call belief polarisation, underneath which are a plurality of phenomena. Next, I will argue that indoctrination is one of these phenomena, i.e. indoctrination can generate belief polarisation. Crucially, it does so when we try to counteract it, thus making the deed even more difficult.

4.3 The polarising effect of indoctrination

In this section, I will argue that indoctrination is part of the group of phenomena— together with Kellyan reasoning, Kripkean dogmatism and partisan cheerleading⁴— which we recognise as instances of belief polarisation on a more surface level. For short, I will say that indoctrination causes or generates belief polarisation. This

⁴I will not discuss here whether echo chambers and sticky beliefs cause belief polarisation in the same way, though intuitively one might think so.

means that if a person who has been indoctrinated exhibits belief polarisation in a debate, having been indoctrinated probably contributes to their belief polarisation. To see how indoctrination can be a cause of belief polarisation, we should first understand belief polarisation in terms of pattern of response to reasons.

In the psychological literature, belief polarisation is typically described as a process through which a *belief*, which is originally held with some degree of strength, is *held* even more firmly as a result of exposure to mixed evidence. But another way to look at the same phenomenon is to focus on how the *believer processes* the mixed evidence: She processes the mixed evidence, in such a way that previously held beliefs are held more strongly than before.

This shift of focus is due to Kelly, who changes the viewpoint when presenting his model and Kripkean dogmatism. From this point of view, it becomes natural to talk about patterns of response to reasons. Belief polarisation can be depicted as a particular function—or a set of functions—that links some features of the environment to our previously held beliefs, such that these features appear to us as reasons to hold our beliefs more strongly. In other words, belief polarisation occurs when we take each piece of the mixed evidence as a reason to increase the confidence in our beliefs.

If we describe belief polarisation this way, we can re-write Kelly's model and Kripkean dogmatism too. Kelly can be seen as presenting the snapshot of how a particular bias affects our responsiveness to reasons. When the mixed evidence comes in, we take the welcome evidence as a reason to increase the confidence in our beliefs; by contrast, we take the unwelcome evidence as a reason to invest more cognitive resources to detect fallacies and to come up with alternative explanations that can make the evidence fit with our beliefs. Kripkean dogmatism can similarly be expressed in terms of pattern of response to reasons. Any evidence that contradicts the dogmatist's belief is taken by her to be a reason to call it misleading information and increase her confidence in her belief.

Now, I have previously assumed that indoctrination is the production of a defective pattern in the reasoning of the subject so that she always chooses to do

the bidding of her manipulator. That defective pattern consists essentially in either dismissing any information that would lead her to stop choosing to do the manipulator's bidding or—if not dismissing—turning that information from undesirable to desirable for the manipulator. Under this assumption, it is quite clear how this defective pattern can lead to polarisation.

The sole *dismissal* of unwelcome evidence may or may not lead to polarisation: Not paying attention to unwelcome evidence may (like in Kripkean dogmatism) or may not be taken as corroborating one's previously held beliefs—that is something to be investigated empirically. But surely, *twisting* the unwelcome evidence, only to interpret it as welcome evidence, is a most powerful push towards polarisation. Quite literally, whatever information is given to the subject of indoctrination, she is going to use it to keep choosing to do the manipulator's bidding: It will not alter her path, but rather reinforce the one she is on already.

To be clear, indoctrination-induced belief polarisation is a unique variety, which can neither be reduced to Kripkean dogmatism, nor Kellyan reasoning, its main competitors. Consider Kripkean dogmatism first. Not only does Kripkean dogmatism *not* predict the *twisting* mechanism, but it also describes a *different dismissal* mechanism. When the Kripkean dogmatist labels the unwelcome evidence as misleading, it is because the content of that evidence logically contradicts the content of the dogmatist's belief. As Kelly puts it: The dogmatist thinks that it follows immediately from her belief that any evidence that contradicts it is misleading. By contrast, the subject of indoctrination does not necessarily think that. What makes the evidence unwelcome for her is that it is inconvenient for the manipulator, not that it is logically inconsistent with her beliefs. Even if the evidence is consistent with her beliefs, she will eventually choose to dismiss it, if that is what the indoctrinator wishes.⁵

⁵This point may raise some questions. I just claimed that the twisting mechanism of indoctrination is not necessarily about the conflict between the beliefs of the subject and evidence that contradicts it, but rather it has to do with the lingering influence of the manipulator's will. The subject tends to explain away information that the manipulator does not want her to have, because the manipulator does not want her to absorb it, not necessarily because it logically contradicts the propositional content of her beliefs. But then, how is the twisting mechanism of indoctrination the same twisting mechanism that we find in echo chambers and sticky beliefs, where there is no such figure as the manipulator casting their influence over the choices of other people? Should we cast doubt, retro-

Indoctrination-induced belief polarisation is also different from Kellyan reasoning. In Kellyan reasoning people typically devote cognitive resources to thoroughly analyse the unwelcome evidence and to find alternative explanations for it. This is a trademark of Kellyan reasoning, but not one of indoctrination. Even if subjects of indoctrination sometimes reason that way, they can also more easily brush off any unwelcome evidence or explain it away with a stock reply, usually provided by the manipulator. In other words, they can resort to less cognitively-costly techniques to neutralise the uncomfortable evidence than the ones described by Kelly.

So, the subject of indoctrination has a distinctive way of becoming polarised in her beliefs. Then, if we assume that more than one mechanism lies underneath the surface-level phenomenon of belief polarisation, we can say that indoctrination is one of them.

4.4 A further harm of indoctrination

Before moving on to discussing solutions for counteracting indoctrination, it is worth noting that linking indoctrination to belief polarisation reveals a further harm of indoctrination. Assuming that indoctrination can be a cause of belief polarisation, then the debate is open on whether it is harmful as such. I am already working under the assumption that indoctrination is undesirable due to its epistemic harmfulness. In addition to that, indoctrination now faces the charge of being pernicious due to its polarising power. A charge, I believe, that is rightfully issued.

At the beginning of this thesis, I argued that indoctrination causes a whole va-

spectively, on the claim that indoctrination, echo chambers and sticky beliefs are all manifestations of the same epistemic phenomenon? No, we should not, because from the viewpoint of doxastic cognition, they are manifestation of the same epistemic phenomenon, i.e. the emergence of the same defective pattern. Independently of how the defective pattern reinforces itself, at some point the protagonists of all three situations—indoctrination, echo chambers and sticky beliefs—find themselves in a mental state that effectively dismisses and explains away systematically any evidence that can potentially change their current mindset, and that can potentially dislodge the defective pattern. So, it is fair to say that indoctrination, echo chambers and sticky beliefs are manifestations of the same epistemic phenomenon, from the viewpoint of doxastic cognition; but is it also fair to say that indoctrination and Kripkean dogmatism are not exactly the same mechanisms, because the Kripkean dogmatist dismisses (and dismisses only) any information that logically contradicts the content of her beliefs, while the subject of indoctrination dismisses and explains away any information that the manipulator—who casts an influence on her choices and epistemic behaviour—does not want her to have.

riety of harms, and I said that I would focus on its epistemic harms. During the dissertation, I have already indicated the instillation of the defective pattern as the distinctive epistemic harm of indoctrination. The reason why I am now interested in addressing this further harm is because it can be directly, closely linked to the defective pattern, as we shall see. Moreover, it will later help me to conclude that indoctrination is a far more central problem in our society than we might have realised. So, I allow myself to briefly deviate from the main argumentation of this chapter and talk about it in this section. My argument goes as follows.

If we want to show that indoctrination-induced belief polarisation is undesirable, the fact that it is a form of belief polarisation is not sufficient, because not all instances of belief polarisation are undesirable. Under most accounts of belief polarisation, there are some instances of it that are deemed unproblematic, rational and even useful to some degree. For instance, under Kelly's account, the mechanism responsible for belief polarisation is essential for the progress of science: Since science is an anomaly-driven process of inquiry, scientists have an imperative to devote more resources to explaining anomalies than regularities. Under Kripkean dogmatism, belief polarisation seems sound when it stems from practical, time-saving, reasonable behaviour such as the one Kripke himself confesses to:

I myself have not read much defending astrology, necromancy, and the like . . . Even when confronted with specific alleged evidence, I have sometimes ignored it although I did not know how to refute it. (Kripke, 2011, p. 23)

If belief polarisation is not undesirable per se, then whether it is undesirable depends on which form of belief polarisation we are considering (Kellyan reasoning, Kripkean dogmatism, etc.), and which particular case we are considering: Non-pernicious cases of belief polarisation are the exception, but they do exist. Above all, we should be suspicious of the claim that belief polarisation is necessarily undesirable in that it is *irrational* because there are a variety of studies that show the opposite; for instance, Singer et al. (2019) argue that “persistent disagreement that grounds political and social polarization can be produced by epistemically rational

agents, when those agents have limited cognitive resources” (p. 2243). So, to show that indoctrination-induced belief polarisation is pernicious, we must be able to link some harm that it causes to the mechanisms that are characteristic of indoctrination and responsible for belief polarisation.

Such a harm, I believe, is the fact that indoctrination does damage to democratic societies from a social and political point of view by *polluting the political discourse*. By political discourse, I mean everything that is said in a society about politics, and especially the conversations—private, public and online—between citizens, what politicians talk about, what the citizens ask of them and what the media broadcast. Further, I am assuming that political discourse can influence how politicians make their decisions and issue laws, especially in a time when massive political movements use the internet to have their voice heard. So, I am also assuming that it is good for democracy that the political discourse is a space of inclusion and dialogue.

Intuitively, there are at least two ways in which the polarising power of indoctrination pollutes political discourse: amplifying the disagreement between the citizens on important matters of fact, and reducing the capacity for dialogue and compromise between them. This intuition is supported by the fact that we can trace both these effects of indoctrination back to the defective pattern.

Disagreement on matters of fact is amplified by indoctrination because the defective pattern, by causing the subjects of indoctrination to select their sources of information arbitrarily, can lead them to believe in arbitrary sets of evidence, and eventually falsehoods. As Kelly explains, we can distinguish between two senses of ‘evidence’. In a narrow sense, evidence “consists of relevant information about the world” (Kelly, 2008, p. 627). In the broad sense it “includes everything of which one is aware that makes a difference to what one is justified in believing” (Kelly, 2008, p. 628). Due to the defective pattern, and if it suits the manipulator, subjects of indoctrination possess broad evidence that is significantly different from the narrow evidence, and also from the broad evidence of the other non-indoctrinated citizens.⁶

⁶Though that is not to say that subjects of indoctrination are the only people who are wrong about matters of fact as a result of defective reasoning, of course.

This means that in society there is, as a result of indoctrination on a large scale, an amplified disagreement not only about various matters of fact of public interest, but also about what evidence is relevant to decide upon those matters. This is a problem in a democratic society, where people are called to vote for policies and participate in the political discourse. When the policymakers and the voters are divided so radically on matters of fact and how to settle them, it becomes very difficult to make policies based on reason and evidence, where facts and evidence should be the grounds for public policies. This is the first way in which indoctrination-induced belief polarisation pollutes the political discourse of democratic societies.

The second way in which it is harmful is that it reduces the overall capacity for dialogue and compromise in society, because the defective pattern causes the subjects of indoctrination to become overconfident in their beliefs, and thus less inclined to engage with people who disagree with them. According to Kelly's *Key Epistemic Fact*:

For a given body of evidence and a given hypothesis that purports to explain that evidence, how confident one should be that the hypothesis is true on the basis of the evidence depends on the space of alternative hypotheses of which one is aware. (Kelly, 2008, p. 620)⁷

According to the Key Epistemic Fact, people should decrease the confidence in their beliefs about a fact when they are given evidence that could explain that fact differently. But the subject of indoctrination, who is affected by the defective pattern and manipulated by a foreign will, is so effective in brushing off and twisting the alternative explanations—often with the aid of stock counter-explanations provided by the manipulator—that we cannot sensibly say that she is aware of any alternatives to her worldview. Consequently, she appears overconfident of her beliefs any time she has an intellectual exchange on a topic her manipulator cares about.

Overconfidence in political discourse can be seen as vicious for many reasons, but the most evident one is that it impacts the overconfident person's capacity of

⁷Similar results which associate polarisation to overconfidence can be found in studies such as Stanley et al. (2019).

listening to other people's opinions, and makes it more difficult for her to change her mind when it is appropriate. Further, overconfidence may easily transform into partisanship, which also is not helpful for the exchange of ideas. Ultimately, this is a problem in society, because indoctrination on a large scale can undermine the citizens' and voters' capacity for dialogue and compromise.⁸

If this is right, then whenever indoctrination produces belief polarisation, it contributes to polluting the political landscape, in that it is partially responsible for deep political disagreement and a reduced capacity for dialogue. All in all, the instillation of the defective pattern produces social, as well as epistemic, harm. This is a further motivation—on top of tackling its epistemic harm—to find a solution to it altogether.

4.5 Two strategies to bypass the defective pattern

If I am right in saying that indoctrination can cause belief polarisation, then counteracting indoctrination becomes particularly tricky.

To frame the kind of task, counteracting indoctrination is *not* a task of debiasing. As Kenyon (2014) notes, “typically three kinds of error may arise in a debiasing attempt: unnecessary correction, overcorrection, and insufficient correction” (p. 2543). Indoctrination is not something the correction of which comes in degrees, mostly because ‘correction’ is not a suitable term for indoctrination. Errors are corrected, biases are corrected, but it is an understatement to call indoctrination an error and a bias. Indoctrination is instead a sophisticated form of manipulation by which a complex mechanism is installed in the subject's mind, and it requires a lot of effort to maintain: It is not innate, it is not a default state of the mind.

This can be seen as good news, because notoriously “debiasing is extraordinarily difficult for the most part. The approaches to it that one might intuitively expect to be effective have an alarming tendency either to be ineffective, or to worsen the bias” (Kenyon, 2014, p. 2543). This also means that we need to find ad hoc measures against indoctrination, independently of the literature on biases.

⁸Again, this is not to say that overconfidence in the political discourse is exclusive of indoctrinated people. Just that indoctrination can be a cause of overconfidence.

As mentioned before, the kind of task we are taking up is to help the subject of indoctrination dislodge her defective pattern of response to reasons. This is a particularly difficult task because the very defect of the pattern is an extra-ordinary resistance to change. Since indoctrination consists in making the subject unable to change the way she thinks, it is difficult for us to undo the work of the manipulator. This is mostly because, as anticipated, the most intuitive approach to help the subject think differently backfires, in that it produces belief polarisation.

The most intuitive strategy, of course, is to try and *talk* the subject of indoctrination out of it. We do this by giving her the evidence that she is wrong about certain matters, or that she has reasons to behave differently from the way she does. For instance, if she is indoctrinated into the cult of a leader, we want to give her proof that he is a charlatan, reasons to stop following him. Sadly, this strategy does not work because reasons are exactly what feeds into the pattern, and are systematically sorted out to the advantage of the manipulator, causing the subject to cling more firmly to her beliefs and to him. So, reasons and persuasion seem to be—as a general rule—to be useless actions to counteract indoctrination.

What seems more promising is a strategy that can somehow *bypass* the defective pattern of response to reasons; a strategy which consists in actions that do not feed the pattern. One such strategy revolves around the notion—mentioned in the previous chapter—of Indoctrination Anxiety. This is a feeling that can be triggered by issuing an Etiological Challenge, i.e. a challenge to reflect whether the causal origin of one's beliefs is contingent and thus in some way unjustified and questionable—e.g. by saying: “You only believe that because you grew up in this particular context!”. Indoctrination Anxiety, more precisely, is the demand:

To try to determine whether one's ability to engage in intellectual self-monitoring might have been compromised all along, by one's having been subjected to a program of indoctrination that has impaired one's ability to fairly assess considerations that tell against one's prior commitments. (DiPaolo and Simpson, 2016, p. 22)

Why should we think that perhaps Indoctrination Anxiety could be able to suc-

ceed where reasoning will not? Because feelings—of which Indoctrination Anxiety is one—are perhaps not entirely subjected to the polarising force of the defective pattern of response to reasons, given that they are not reasons. Even after we rationalise our feelings, they may persist; they cannot be dismissed or explained away as information can.

To be sure, feelings are often taken to be reasons for acting in a particular way. For example, one may take the feeling of guilt as a reason not to do something or to ask for forgiveness. And feelings can even be processed in a twisted way. For instance, a person indoctrinated into religious beliefs can take the feeling of doubt not as a reason to examine her beliefs, but as a reason to believe that God is testing her faith, out of love. But even when a feeling is rationalised this way, it does not necessarily go away. The faithful may tell herself (or may be told) that doubt is a sign that God is looking after her, but the feeling of doubt may nevertheless persist.

Indoctrination Anxiety may be rationalised by the subject of indoctrination. When challenged to think whether she only believes what she does because of indoctrination, she may be able to explain how that is false, or even true but not a problem—because the defective pattern of response to reasons causes her to. But the anxiety may keep looming in the background, and thus represents a threat to the defective pattern, in that it may eventually lead the subject of indoctrination to realise that she has indeed been indoctrinated and it is a problem.

But inducing Indoctrination Anxiety is not the only promising strategy for bypassing the defective reasoning of the subject of indoctrination. There is a second way of doing so. That is by asking a simple question. Before the subject of indoctrination is presented with mixed evidence and is drawn to radicalisation by the polarising force of indoctrination, we can ask her: What *could* make you change your mind?

The first key idea behind this proposal is a consideration of *how* the subject seems to maintain her beliefs—the ones that the manipulator wishes her to have. These beliefs are not unfalsifiable, but they are maintained in a way that they de facto are unfalsifiable in her eyes. If any evidence that could potentially falsify these

beliefs is systematically dismissed or explained away, then it is virtually equivalent for the subject to maintain an unfalsifiable belief or a falsifiable indoctrinated belief.

For example, the belief that the world is going to end on a precise day is a falsifiable belief, and beliefs of this kind have in fact been falsified in the past. But famously—a whole literature on cognitive dissonance deals with this fact—this does not seem to affect the members of apocalyptic cults, who only do minor adjustments to their beliefs in the face of falsifying evidence. Their beliefs remain unfalsified in their eyes, and so does the net of beliefs that are interconnected with them.

The second key idea behind the proposal is inspired by DiPaolo's (2020) suggestion to draw inspiration from *converts* for ways to narrow the gap between opposite beliefs in the context of belief polarisation. He argues:

My interest in converts is based in optimism that we can get clues on how to reduce divisions in belief by studying their intellectual transformations. But in order to ensure appropriate expectations I want to critically examine converts' epistemic credentials. (DiPaolo, 2020, p. 93)

In his analysis, DiPaolo highlights a limit of converts. He shows that citing their conversion is not always sufficient for winning an Etiological Challenge, if the Etiological Challenge is aimed at criticising the current closed-mindedness of the convert, and if the convert is indeed closed-minded. Additionally, he shows that conversion itself can incentivise closed-mindedness, because it can lead the convert to succumb to epistemic self-licencing, i.e. believe that they are more epistemically virtuous than they are, and thus be less epistemically vigilant.

Nevertheless, the notion of radical intellectual transformation, i.e. conversion, is indeed a powerful one in the context of indoctrination and belief polarisation, and so is the concept of convert, i.e. someone who has gone through conversion in life, and who most likely knows what it took them to radically change their worldview.

We can expect to be able to appeal to this awareness, and receive sensible answers, when we ask converts: What could make you change your mind again?

What sort of evidence, what sort of experience, would hypothetically cause you to change your mind about your current beliefs once again? What could cast the doubt on your convictions right now?

Assuming that appealing to a convert's conversion is useful, this idea can be tweaked and imported in the context of indoctrination, in that we can ask the subject—before she is given the evidence against her current beliefs—what sort of evidence she can anticipate to be able to shift her opinion about the subject matter. The hope is that this challenge will be received without being processed by the defective pattern and that it will lead the subject to think outside the box the manipulator has set up for her.

Asking the subject what could change her mind is a counterfactual strategy, so to speak, and it is effective because it clashes with the apparent unfalsifiability of the indoctrinated beliefs. Once the subject realises that nothing *could* change her mind, she might start thinking about what *should* be able to change her mind. She might also become more sensitive to those reasons, and she might acquire a feeling of doubt concerning her beliefs and epistemological practices. Achieving these feats is the sign that the greatest obstacle to counteracting indoctrination has been overcome.

Consider the following, final example. A subject of indoctrination is under the influence of a cult leader. She is presented with evidence that is uncomfortable for him, so she is caused by the defective pattern to neutralise the evidence. Now, if my speculation is correct, to counteract indoctrination we can ask the subject before the mixed evidence is presented to her: What could make you change your mind about this issue?

There are many things the subject of indoctrination can reply to such a challenge. Perhaps, the manipulator is so effective, as to have prepped her for such a challenge, and explained to her—to the manipulator's benefit—what exactly should count as mind-changing evidence. But if the manipulator has not prepped her for such a challenge, then the subject of indoctrination might surprise herself as replying that nothing could change her mind, which should sound like a suspicious

answer to whoever cares about grounding their beliefs in evidence and facts.

Alternatively, the subject of indoctrination can even more dramatically realise that what would make her change her mind would be some truth about her cult leader: that the leader is not reliable, that the leader has an interest in the subject maintaining her current belief, or in the subject altering her belief. In other words, she could realise that the status of her belief is dependent on the interests of the manipulator and on her relationship with him. This could be a wake-up call for the subject, who could eventually put her beliefs in a new light.

As the example shows, challenging the subject this way is a promising strategy, and there are two ways to try and capture why that is. First, similarly to Indoctrination Anxiety, this challenge seems to bypass the defective pattern by targeting the feelings of the subject, rather than her thoughts: It instils doubt and it sensitises her to reasons that she would otherwise not pay attention to. Second, it forces her to focus on her epistemic practices, her beliefs about what are the best epistemic practices, and whose agenda her epistemic practices benefit. In other words, by challenging her, instead of just giving her another opportunity to think *by way of* the defective pattern, we make her reflect *on* that pattern.

In this chapter, I argued that my epistemological account of indoctrination compels us to find ways to counteract indoctrination by dislodging the defective pattern from the mind of the subjects. I argued that the most obvious strategy to do so—giving the subject evidence that is meant to change her beliefs and mindset—backfires because it generates belief polarisation, i.e. it strengthens the defective pattern instead of weakening it. To do so, I had to take a step back to show that belief polarisation is a surface level phenomenon, underneath which lies a plurality of phenomena, including Kellyan reasoning, Kripkean dogmatism and partisan cheer-leading. I showed that indoctrination belongs to this group, and as such it causes a social harm alongside its characteristic epistemic harm. I finally argued that there are two promising strategies to dislodge the defective pattern—which can only be done if we somehow bypass it. First, we can cause Indoctrination Anxiety in the subject. Second, we can ask her what could possibly change her mind about her

indoctrinated beliefs.

My diagnosis and suggested remedy to indoctrination, in this chapter, relies upon the output of my descriptive inquiry in the previous two chapters, in that it shows that we cannot counteract indoctrination by re-educating the subject, since indoctrination was never about bad education to begin with. A subject of indoctrination is not necessarily someone who has received improper education—she is someone who finds herself in a particular mental state, who reasons in a defective manner as a result of manipulation.

We are tempted to help the subject reason her way out of this mental state by giving her evidence that she should be making different choices and revise her beliefs. But this is not a successful strategy because we are assuming that what it means to be indoctrinated is to have had beliefs inculcated in you by a teacher during the formative years. We cannot help the subject challenge the bad education that she received, because bad education is not what she received; instead, she has been manipulated to the point where the defective pattern of reasoning took root in her mind.

Attempting to reason with the subject is useless because it overlooks the fact that the reasoning pattern of the subject explains away the reasons she has for revising her beliefs and epistemic practices. My suggestion to aim for the subject's feelings or some higher-order self-awareness thus relies upon having dropped the assumption that indoctrination is a degenerate form of education and having assumed that it is the exploitation of an epistemic-cum-cognitive vulnerability of the mind which—instead of inculcating beliefs—instils in the subject a defective way of reasoning.

Concluding remarks

The overarching goal of this work was to show that developing a descriptive analysis of indoctrination from an epistemological point of view is a promising strategy for answering the question: What is indoctrination? By developing my account of indoctrination along these lines, I hope to have delivered on the implicit promise that the proof is in the pudding.

What I take to be the sign of having conducted a successful analysis of indoctrination is that my account explains at various points why previous attempts in the literature had been unsatisfying. First, my account explains why the strategy of defining indoctrination as a form of bad education—education in which (i) the educator has a wicked motivation, or where (ii) the beliefs that are instilled are false or unwarranted, or where (iii) the method of teaching is coercive, cruel, or disrespectful of the pupil's rationality—was never going to succeed.

Thinking of indoctrination as bad education is tempting because there is indeed a connection between education and indoctrination: Indoctrination is sometimes carried out in schools. However, this connection is only superficial, and it is misleading because it overshadows a phenomenon that serves as a better referent to our notion of indoctrination. According to my analysis, what we should be referring to when thinking about indoctrination is not bad education, but rather the exploitation of an epistemological-cum-cognitive vulnerability.

Our mind is vulnerable to the emergence of a defective pattern of reasoning—a self-reinforcing pattern that makes us insensitive to relevant evidence—and indoctrination occurs whenever a manipulator actively seeks to instil this pattern in his subject because it suits him. The manipulator may be a teacher trying to instil

the pattern in his pupils, but he may alternatively be a political leader, a spiritual guide, a parent, or another authority figure. He may not only be indoctrinating his subjects in an education setting, but he may also be doing so through political propaganda, mystical rituals, and so on. So, indoctrination is best understood as a form of manipulation—as the exploitation of a precise vulnerability of the mind to the indoctrinator’s advantage—and not as bad education, which is why the traditional strategy for explicating it was short-lived.

Second, my account explains why trying to separate indoctrination from proper education was so difficult for authors in the literature. According to the literature, there is a trajectory that proper education must follow, and some disturbing factors that sometimes deviate it. Accounting for indoctrination this way meant recognising those disturbing factors, and explicating the difference between indoctrination and proper education. But I have argued that this formula does not work because indoctrination occurs independently of any particular method, teachings or intentions of an educator; rather, indoctrination occurs whenever a subject ends up in a particular mental state, namely the defective pattern. And this mental state does not exclusively arise as a result of improper education.

When we misconstrue indoctrination, we may think it belongs to the class of bad education practices, together, for instance, with sloppy education—a type of education where the teacher does his job in an excessively casual, careless and un-systematic manner. But classifying indoctrination this way does not explain why we do not find other bad education practices as morally objectionable as indoctrination. A more appropriate and striking way to classify indoctrination is to group it with other situations where a person ends up having the same defective pattern of reasoning, for instance, echo chambers and sticky beliefs.

That is why drawing the line between indoctrination and proper education is so difficult: We are not carving reality at its joints, so to speak; so, even if we draw that line, it is not very significant. If we want to carve reality at its joints, mark out different phenomena in a more illuminating way, we ought to separate indoctrination from these other phenomena—echo chambers, sticky beliefs, and perhaps

more—which like indoctrination cause the emergence of the defective pattern.

Third, my account shows why we cannot counteract indoctrination with facts and figures. Under the assumption that indoctrination is bad education, we should be able to contrast it by supplementing it with proper education—by (i) putting a well-meaning teacher in charge, (ii) devising a truthful and liberal curriculum, and (iii) employing teaching techniques that respect the pupils as persons and rational agents. But I have shown that once someone is indoctrinated, giving them reasons to revise what they learnt through what we erroneously call bad education is useless and it can even backfire, in that it can generate belief polarisation.

That is because what the subject of indoctrination learns through bad education is not a set of beliefs that call for being disproved, but rather a defective way of thinking, one which is difficult to alter—precisely because this pattern makes the subject insensitive to any evidence that could alter it. Indoctrination was never so much about bad education and more about pushing the subject towards acquiring a defective and self-reinforcing way of reasoning. Hence my suggestion not to try and persuade the subject out of her defective pattern with facts and evidence—that is, by re-educating her—and rather aim at eliciting feelings of anxiety and doubt about her epistemic condition, and force her to become self-aware of her epistemic practices.

Finally, my account explains why indoctrination is a far more central concept in our conceptual repertoire—and a more compelling problem—than we might have realised. Because indoctrination is not reducible to bad education, we can see a crucial reason why there are so many bad epistemic outcomes in otherwise well-educated, liberal societies. That is, our epistemic-cum-cognitive vulnerability that leads to the emergence of the defective pattern can be exploited even if the subject receives proper education in school.

Further, while the exploitation of this vulnerability is an epistemic harm per se, the emergence of the defective pattern does damage to democratic societies by polluting the political discourse, and more precisely by amplifying the disagreement between the citizens on important matters of fact, and reducing the overall capacity

for dialogue and compromise.

Combine this with the notion that indoctrinators have at their disposal echo chambers and sticky beliefs as tools to strengthen the defective pattern of their subjects, and it becomes clear that indoctrination is not a singular, isolated matter, a problem that only pertains education in insulated communities and authoritarian regimes. Rather, it can permeate and affect every part of our society.

Conducting a descriptive analysis, as opposed to a conceptual one, which is what authors in the tradition do, has allowed me to drop some of the common assumptions about indoctrination, and to investigate the phenomenon without worrying too much about upsetting our preconceptions. As I have shown, important aspects of indoctrination come to light, once we stop trying to understand *our* concept of indoctrination—the fuzzy idea that it is a degenerate form of education—and we start trying to understand the actual phenomenon of indoctrination, with the aid of science where possible.

Because my account relies on empirical findings in Psychology—for instance the literature on just world beliefs and belief polarisation—and lends itself to further research that could benefit from empirical investigation in Psychology and cognitive science—perfectly in line with Haslanger’s definition of what constitutes a descriptive analysis—one could see my project as being conducted in the spirit of *Naturalised Epistemology*.

In his essay on the topic, Quine (1969), disillusioned with the project of traditional Epistemology—which aims to reduce all knowledge, including the natural sciences and Mathematics, to simpler, self-evident truths—claims that:

Epistemology still goes on, though in a new setting and a clarified status. Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a human subject. (Quine, 1969, p. 82)

In this naturalised form, Epistemology is the study of how, from a psychological point of view, humans process physical stimulation into a description of the external world. It analyses how we transform sensory evidence into beliefs and

knowledge. As a result, there is a shift: While classical Epistemology aspired to contain the natural sciences, Naturalised Epistemology is contained in the natural sciences, as a chapter of Psychology.

This move can be taken to legitimise the aid of empirical methods in theorising how we build knowledge. As Wrenn (2003) puts it:

Naturalistic epistemology is an approach to the theory of knowledge that emphasizes the application of methods, results, and theories from the empirical sciences. It contrasts with approaches that emphasize a priori conceptual analysis or insist on a theory of knowledge that is independent of the particular scientific details of how mind-brains work. (Wrenn, 2003)

It is in this spirit that one can incorporate empirical methods and empirical findings in the epistemological account of a particular phenomenon, such as indoctrination. In developing the epistemology of indoctrination, we need not derive our definition of indoctrination from simpler, more familiar concepts, or a set of self-evident assumptions. Instead, we can benefit from the aid of years of empirical research in Psychology and cognitive science, because there is a legitimate interplay between the field of Epistemology and the natural sciences.

As a concluding remark, although I take my analysis to be a descriptive one, I hope it has an ameliorative aspect to it as well. When we conduct a descriptive analysis, according to Haslanger, we gain new knowledge about a kind, which serves to update our concept of that kind. But there is something intrinsically valuable to updating our concepts, and especially concepts that play an important role in society, such as indoctrination.

Indoctrination is a misused term and a misunderstood phenomenon. In popular culture, we use the term ‘indoctrination’ as an insult to dismiss practices we do not like by appealing to a notion of brainwashing that is borrowed from dystopian literature and toxic political rhetoric. In Philosophy of Education, we have failed to explain the most relevant link between indoctrination and proper education.

By contrast, in this work, I tried to show that our term ‘indoctrination’ can be

taken to refer to a more interesting phenomenon—the exploitation of a cognitive-cum-epistemological vulnerability. This, I argued, is the phenomenon that lies in the most proximate vicinity of what we take to be saying when we talk about indoctrination—not so-called brainwash, and not a degenerate form of education.

So, ‘indoctrination’ is a word that currently comes attached with multiple meanings. This obscures a phenomenon that we have an interest in understanding and counteracting because it is epistemologically and politically harmful. The present analysis is hopefully ameliorative because attending to our concept of indoctrination by offering a more accurate, technical account of it can potentially help demystify indoctrination and address it more effectively and respectfully.

Bibliography

- Ahlstrom, K. (2008). Epistemology and Empirical Investigation. *Grazer Philosophische Studien*, 76(1):109–134. Editions Rodopi.
- Atkinson, R. F. (2010). Indoctrination and Moral Education. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 43–51. Routledge.
- Chignell, A. (2018). The Ethics of Belief. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2018 edition. <https://plato.stanford.edu/archives/spr2018/entries/ethics-belief/>. Last accessed on December 9, 2020.
- Crittenden, B. S. (2010). Indoctrination as Mis-Education. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 102–117. Routledge.
- Dentith, M. R. X. (2019). Conspiracy theories on the basis of the evidence. *Synthese*, 196(6):2243–2261.
- DiPaolo, J. (2020). Conversion, Causes, and Closed-Mindedness. *Journal of the American Philosophical Association*, 6(1):74–95.
- DiPaolo, J. and Simpson, R. M. (2016). Indoctrination Anxiety and the Etiology of Belief. *Synthese*, 193(10):3079–3098. Springer Nature.
- Feinberg, J. (2007). The child's right to an open future. In Curren, R. R., editor, *Philosophy of Education: An Anthology*. Blackwell.

- Feldman, R. and Conee, E. (2001). Internalism Defended. *American Philosophical Quarterly*, 38(1):1–18. University of Illinois Press.
- Flew, A. (2010a). Indoctrination and Doctrines. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 52–71. Routledge.
- Flew, A. (2010b). Indoctrination and Religion. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 82–90. Routledge.
- Garnett, M. (2015). Freedom and Indoctrination. *Proceedings of the Aristotelian Society*, 115(2pt2):93–108. Wiley.
- Garrison, J. W. (1986). The paradox of indoctrination: A solution. *Synthese*, 68(2):261 – 273. Springer.
- Gatchel, R. H. (2010). The Evolution of the Concept. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 7–13. Routledge.
- Green, T. F. (2010). Indoctrination and Beliefs. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 20–36. Routledge.
- Gregory, I. M. M. and Woods, R. G. (2010). Indoctrination: Inculcating Doctrines. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 126–146. Routledge.
- Hannon, M. (forthcoming). Disagreement or badmouthing? the role of expressive discourse in politics. In Edenberg, E. and Hannon, M., editors, *Political Epistemology*. Oxford: Oxford University Press.
- Haslanger, S. (2000). What Are We Talking About? The Semantics and Politics of Social Kinds. *Hypatia*, 20(4):10–26. Wiley Online Library.

- Kasperbauer, T. J. (2015). Psychological Constraints on Egalitarianism: The Challenge of Just World Beliefs. *Res Publica*, 21(3):217–234. Springer Netherlands.
- Kelly, T. (2008). Disagreement, Dogmatism, and Belief Polarization. *Journal of Philosophy*, 105(10):611–633. Journal of Philosophy Inc.
- Kenyon, T. (2014). False polarization: debiasing as applied social epistemology. *Synthese*, 191(11):2529–2547. Springer Verlag.
- Kilpatrick, W. H. (2010). Indoctrination and Respect for Persons. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 37–42. Routledge.
- Kripke, S. (2011). On Two Paradoxes of Knowledge. In Kripke, S. A., editor, *Philosophical Troubles. Collected Papers Vol I*. Oxford University Press.
- Mahoney, J. (2014). Justice in education and religious freedom. *Social Philosophy and Policy*, 31(1):276–294.
- Mills, C. (2003). The child's right to an open future? *Journal of Social Philosophy*, 34(4):499–509. Wiley-Blackwell.
- Montague, P. (2000). The Myth of Parental Rights. *Social Theory and Practice*, 26(1):47–68. Department of Philosophy, Florida State University.
- Moore, W. (2010). Indoctrination and Democratic Method. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 72–77. Routledge.
- Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2):141–161.
- Overall, C. (2010). Indirect Indoctrination, Internalized Religion, and Parental Responsibility. In Caws, P. and Jones, S., editors, *Religious Upbringing and the Costs of Freedom: Personal and Philosophical Essays*. Pennsylvania State University Press.

- Quine, W. V. O. (1969). Epistemology Naturalized. In *Ontological Relativity and Other Essays*, pages 69–90. Columbia University Press.
- Shiffrin, S. V. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy and Public Affairs*, 29(3):205–250. Wiley-Blackwell.
- Singer, D. J., Bramson, A., Grim, P., Holman, B., Jung, J., Kovaka, K., Ranginani, A., and Berger, W. J. (2019). Rational social and political polarization. *Philosophical Studies*, 176(9):2243–2267. Springer Verlag.
- Snook, I. A. (2010a). *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*. Routledge.
- Snook, I. A. (2010b). Indoctrination and Moral Responsibility. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 118–125. Routledge.
- Sripada, C. S. (2012). What Makes a Manipulated Agent Unfree? *Philosophy and Phenomenological Research*, 85(3):563–593. Wiley-Blackwell.
- White, J. P. (2010). Indoctrination and Intentions. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 91–101. Routledge.
- Wilson, J. (2010). Indoctrination and Rationality. In *Concepts of Indoctrination (International Library of the Philosophy of Education Volume 20): Philosophical Essays*, pages 14–19. Routledge.
- Wrenn, C. B. (2003). Naturalistic epistemology. *Internet Encyclopedia of Philosophy*. <https://iep.utm.edu/nat-epis/>. Last accessed on December 9, 2020.
- Yaffe, G. (2003). Indoctrination, coercion and freedom of will. *Philosophy and Phenomenological Research*, 67(2):335–356. Wiley-Blackwell.